

Domingos Fernandes

AVALIAÇÃO DAS APRENDIZAGENS: UMA AGENDA, MUITOS DESAFIOS



Texto Editora
www.TE.pt

**EVOLUÇÃO
E QUALIDADE**

ÍNDICE

Introdução	3
Algumas Referências Recentes	5
Três Razões Suficientes para Mudar a Avaliação	7
Primeira Razão: Aprendizagem	7
Segunda Razão: Currículo	7
Terceira Razão: Democracia	8
Conceptualizações de Avaliação	10
A Avaliação como Medida	10
A Avaliação como Descrição	11
A Avaliação como Juízo de Valor	11
Síntese e Limitações das Três Gerações de Avaliação	12
A Avaliação como Negociação e como Construção	13
Avaliação Psicométrica e Avaliação Alternativa	14
Princípios e Orientações para uma Avaliação Alternativa	16
Integração Ensino-Aprendizagem-Avaliação	16
Seleção de Tarefas	16
Funções	18
Triangulação	18
Triangulação de estratégias, técnicas e instrumentos	19
Triangulação de intervenientes	19
Triangulação de espaços e de tempos	19
Transparência	19
A Essencialidade do <i>Feedback</i> na Avaliação Alternativa	20
O <i>Portfolio</i> como Exemplo de Avaliação Alternativa	21
Os Professores e a Avaliação das Aprendizagens	23
Avaliação Externa	25
Exames Públicos Nacionais	25
Breve bosquejo histórico dos exames	26
Características gerais dos exames	26
Funções dos exames	28
Considerações sobre o tipo de questões	30
Equidade	32
Validade e fiabilidade dos exames	33
Vantagens e desvantagens dos exames	35
Avaliação Aferida	36
Estudos Internacionais de Avaliação das Aprendizagens	42
Uma Agenda, Muitos Desafios	48
Referências	52

INTRODUÇÃO

Este texto* surge numa altura em que estão em curso, ou estão em vias de concretização, um conjunto de medidas de política relativas ao sistema educativo português. Entre outras, podem destacar-se a gestão flexível do currículo na educação básica, o início de uma revisão curricular abrangendo percursos de educação e de formação de nível secundário, a aplicação de provas de aferição nos anos terminais dos ciclos da educação básica, a introdução de exames nacionais no final da escolaridade básica, a passagem para a tutela do Ministério da Educação de toda a formação inicial de natureza vocacional ou profissionalizante ou o alargamento da escolaridade obrigatória para 12 anos. Por outro lado, como é sabido, prepara-se uma revisão da Lei de Bases do Sistema Educativo que, supostamente, irá enquadrar, *a posteriori*, decisões de política educativa publicamente anunciadas pelo Ministro da Educação ou já sob a forma de normativo legal como é o caso de algumas das que acima se referem.

No que concerne à avaliação das aprendizagens dos alunos do ensino não superior, estamos num momento particularmente importante para fazer um ponto de situação tão desapassionado e imparcial quanto possível, acerca dos problemas, dos desafios, das potencialidades ou das questões relativas ao seu desenvolvimento. O mesmo é dizer que interessa analisar o que se passa actualmente com a avaliação das aprendizagens, nas suas múltiplas e complexas dimensões, para que se possam identificar linhas de acção que contribuam para democratizar, modernizar e melhorar o sistema educativo.

Importa antes do mais referir que a avaliação das aprendizagens é aqui entendida como todo e qualquer processo deliberado e sistemático de recolha de informação, mais ou menos participado, mais ou menos negociado, mais ou menos contextualizado, acerca do que os alunos sabem e são capazes de fazer numa diversidade de situações. Normalmente este processo permite a formulação de apreciações por parte dos diferentes intervenientes (incluindo os próprios alunos), acerca do mérito ou valor do trabalho desenvolvido pelos alunos o que, em última análise, permitirá tomar decisões que regulem os processos de aprendizagem e de ensino. Assim, neste sentido amplo, a avaliação das aprendizagens inclui a avaliação de conhecimentos, de desempenhos, de capacidades, de atitudes, de procedimentos ou de processos mais ou menos complexos de pensamento. Se quisermos, inclui a avaliação de competências, ou a avaliação dos saberes em utilização, como agora se ouve com alguma frequência, a partir de concepções dominantes nos meios económicos e empresariais. (Por razões de simplificação de escrita, sempre que neste texto surgir a palavra *avaliação* e se nada for dito em contrário, estarei a referir-me à avaliação das aprendizagens.)

Voltando então ao raciocínio que antes seguia, poder-se-á dizer que a avaliação, componente indissociável do ciclo constituído pelo ensino e pela aprendizagem, constitui um elemento essencial de desenvolvimento dos sistemas educativos porque é através dela que, por exemplo, as escolas podem empobrecer ou enriquecer o currículo, os professores podem organizar o ensino com maior ou menor ênfase na experimentação ou na resolução de problemas, os alunos podem estudar com maior ou menor motivação, os pais e encarregados de educação podem acompanhar a vida escolar dos seus filhos ou educandos com maior ou menor interesse, a sociedade em geral pode estar mais ou menos satisfeita com o que os jovens estão a aprender ou que os governos podem, ou não, decidir afectar recursos a um dado projecto, a uma dada escola ou a um conjunto de escolas.

* Este texto é uma versão não definitiva de um trabalho em desenvolvimento para publicação posterior na Texto Editora.

Repare-se que, em princípio, governos, políticos, escolas, gestores escolares, professores, pais e alunos estão todos interessados na avaliação, precisam dela ou utilizam-na, mais ou menos sistematicamente, de diversas formas. Por exemplo, os governos e os políticos precisam de avaliar para estabelecer *standards* (padrões de excelência se quisermos) para monitorizar a qualidade da educação, para perceber os efeitos dos exames nacionais ou para formular políticas. As escolas e os respectivos gestores utilizam a avaliação para identificar pontos fortes e pontos fracos dos seus projectos educativos, para planificar ou melhorar projectos e programas em curso ou para intervir na gestão dos recursos humanos e materiais. Os professores e os diversos responsáveis pela vida das escolas utilizam a avaliação para monitorizar o progresso dos alunos, para avaliar o currículo e proceder ao seu refinamento, para introduzir correcções no processo de ensino, para motivar os alunos ou para preparar a atribuição de classificações. Finalmente, os pais e os alunos podem utilizar a avaliação para ajuizar acerca do trabalho realizado pelas escolas, para tomar decisões conscientes relativas a prosseguimento de estudos, para analisar pontos fortes e pontos fracos dos alunos ou para regular o processo de aprendizagem.

Assim, num contexto de alterações curriculares e tendo em conta o que acima se refere, parece fazer sentido que neste texto se procurem equacionar e sistematizar conceitos, ideias ou procedimentos relacionados quer com a **avaliação interna** (avaliação desenvolvida e integralmente controlada pelos professores e pelas escolas) quer com a **avaliação externa** (avaliação desenvolvida e normalmente totalmente controlada por uma instituição ou instituições externas à escola; no nosso caso, é normalmente o Ministério da Educação através de alguns dos seus departamentos, gabinetes ou direcções gerais). No que se refere à avaliação interna analisam-se as principais características da avaliação alternativa, também designada por avaliação autêntica, avaliação contextualizada ou avaliação educativa, em contraste com as características da avaliação psicométrica, fazem-se considerações relativas a práticas de professores identificadas por algumas investigações realizadas em Portugal, identificam-se e discutem-se as suas principais funções, analisam-se e discutem-se as suas relações com o ensino e com a aprendizagem e referem-se estratégias e instrumentos de avaliação. Quanto à avaliação externa, analisam-se e discutem-se várias questões relacionadas com os exames nacionais do ensino secundário, nomeadamente no que se refere aos seus possíveis efeitos no desenvolvimento do currículo, à sua validade e fiabilidade e às suas vantagens e desvantagens. No que se refere às provas aferidas, faz-se uma análise diacrónica do seu desenvolvimento entre nós, procurando caracterizar as razões que parecem justificar o seu, pelo menos aparente, fracasso.

Portugal participou e continua a participar em estudos internacionais de avaliação promovidos e orientados quer pela *International Association for the Evaluation of Educational Achievement* – IEA – quer pela Organização para a Cooperação e Desenvolvimento Económico – OCDE. No primeiro caso o nosso país participou no *Third International Mathematics and Science Study* (TIMSS) e, no segundo, participou e continua a participar no *Programme for International Student Assessment* (PISA). Nestas condições e considerando o crescente impacto destes estudos de avaliação externa na formulação das políticas educativas dos países participantes, pareceu-me interessante tecer algumas considerações acerca da natureza destes estudos, dos seus propósitos e dos seus impactos.

Um outro objectivo deste texto é o de propor uma agenda para a acção no domínio da avaliação das aprendizagens. Refira-se que é **uma** agenda e não **a** agenda! Ou seja, é uma agenda que resulta duma leitura que faço do actual estado das coisas e que «interfere» no domínio das políticas educativas, no domínio da organização e funcionamento das práticas pedagógicas e didácticas das escolas, no domínio da formação de professores e no domínio da produção de conhecimento por parte das instituições do ensino superior. É evidente que podem surgir outras agendas, reflectindo outras visões, outras concepções e, consequentemente, outras prioridades. O que sinceramente espero é que a agenda proposta contribua para que, no domínio da avaliação das aprendizagens, possamos conjugar esforços da formação, da investigação e das práticas que se traduzam em linhas de orientação devidamente fundamentadas que ajudem a consolidar o que tem funcionado bem, a melhorar o que tem funcionado menos bem e a pôr a funcionar o que, muito simplesmente, não tem funcionado.

ALGUMAS REFERÊNCIAS RECENTES

Nos finais dos anos 80 e muito particularmente nos anos 90 do passado século, assistiu-se a uma grande proliferação de reflexões, investigações e considerações de natureza teórica e prática relativamente à avaliação das aprendizagens dos alunos. Instituições e autores de insuspeita credibilidade produziram, e continuam a produzir, milhares de páginas onde se discutem questões de desenvolvimento dos sistemas educativos, do ensino e das aprendizagens e os papéis que a avaliação tem desempenhado e deverá vir a desempenhar nesse desenvolvimento. Pareceu-me interessante referir aqui algumas publicações que considero mais relevantes pela notória influência que, um pouco por todo o mundo, têm exercido junto de investigadores, educadores e professores, particularmente interessados na avaliação, junto de responsáveis por instituições que trabalham em avaliação ou mesmo junto de responsáveis políticos. Trata-se apenas de uma sugestão de um conjunto de referências que conheço bem e não de uma lista exaustiva e sistemática de todas as publicações de qualidade que surgiram no período em referência. **(Publicações resultantes de reflexões e de investigações realizadas entre nós são referidas, de forma mais contextualizada, ao longo do texto.)**

Ainda nos finais dos anos 80, Desmond Nuttall, investigador inglês prematuramente falecido, editou um livro notável integralmente dedicado à avaliação das aprendizagens com um conjunto de artigos que abordam questões que vão desde a avaliação em larga escala, tal como se organiza em vários países do mundo, até às questões relacionadas com a procura de alternativas à avaliação predominante nos sistemas educativos (Nuttall, 1986). Nos inícios dos anos 90 é publicado um conjunto de livros de referência incontornável no domínio da avaliação das aprendizagens (Berlak, Newmann, Adams, Archbald, Burgess, Raven e Romberg, 1992; Gifford e O'Connor, 1992; Gipps, 1994; Stake 1991a, 1991b). Todos eles põem uma particular ênfase, entre outros assuntos, nos paradigmas que, digamos assim, sustentam diferentes abordagens teóricas e práticas de avaliação, na organização e funcionamento de exames nacionais e de avaliações em larga escala sem qualquer impacto na progressão dos alunos, nas questões de validade, fiabilidade, equidade e comparabilidade dos instrumentos utilizados e/ou dos resultados por eles produzidos ou ainda nas políticas de avaliação. Permito-me destacar o excelente conjunto de artigos constantes nos dois volumes editados por Stake (1991a, 1991b). Num deles discutem-se as relações entre as necessidades de reformar a educação e as políticas de avaliação. São apresentadas e discutidas perspectivas sobre políticas de avaliação em diferentes países, questões referentes à comparação entre escolas a partir da utilização dos resultados em provas de avaliação externa e formas de lidar com as diferenças e de preservar a individualidade. No outro analisam-se os efeitos da avaliação externa obrigatória (com ou sem efeitos na progressão dos alunos) nas práticas de ensino. Começam por analisar-se os efeitos positivos e negativos da avaliação externa na organização e funcionamento das escolas, depois apresentam-se, discutem-se e analisam-se percepções de professores sobre o impacto das avaliações externas no ensino e, mais geralmente, nas salas de aula e, finalmente, em três artigos, ilustra-se como é que questões de natureza conceptual se relacionam com mudanças nas políticas de avaliação.

É ainda no início dos anos 90 que a Associação de Professores de Matemática dos Estados Unidos da América, o *National Council of Teachers of Mathematics* (NCTM), à imagem do que aconteceu com outras instituições e associações de professores de outras disciplinas, dedica inteiramente o seu *Yearbook* de 1993 à avaliação das aprendizagens nas salas de aula de matemática (Webb e Coxford, 1993) e, dois anos passados, publica um documento programático de grande relevância – *Assessment Standards for School Mathematics* (NCTM, 1995) – que estabelece um conjunto de *standards* destinados a orientar as práticas de avaliação dos professores de matemática. Este último é um documento cujo conteúdo é perfeitamente transferível para outros contextos (disciplinares ou não).

Após a primeira metade da década de 90, a *Association of Teacher Educators* norte-americana publicou o seu *Handbook of Research on Teacher Education* (Sikula, 1996) no qual se inclui um artigo dedicado à avaliação autêntica (Tellez, 1996). Merece ainda referência um livro editado por Gary Phye que inclui um conjunto de 18 artigos de natureza teórica e prática (avaliação das aprendizagens nos jardins de infância e nas aulas de matemática, de estudos sociais, de línguas, de artes visuais) e artigos relativos à avaliação em larga escala e ao desenvolvimento de *standards* (Phye, 1997).

Mais recentemente, a prestigiada editora holandesa *Kluwer Academic Publishers* publicou a segunda edição do clássico *Evaluation Models: View Points on Educational and Human Services Evaluation* (Stufflebeam, Madaus e Kellaghan, 2000), cuja primeira edição data de 1983. Nesta segunda edição os editores mantiveram apenas sete artigos da edição anterior, promoveram a revisão de três e acrescentaram quinze. Apesar de se

centrar essencialmente na avaliação de programas, o livro contém artigos de natureza teórica relevantes para a compreensão e conceptualização da avaliação das aprendizagens (Madaus e Kellaghan, 2000; Madaus e Stufflebeam, 2000; Scriven, 2000; Stake, 2000). Inclui ainda dois artigos mais directamente relacionados com a avaliação das aprendizagens (Madaus, Haney e Kreitzer, 2000; Kellaghan e Madaus, 2000).

A *American Educational Research Association* (AERA), publicou a quarta edição do seu *Handbook of Research on Teaching* (Richardson, 2001) na qual se inclui um artigo sobre o papel da avaliação na sala de aula, no ensino e na aprendizagem (Shepard, 2001). Refira-se que este *Handbook* é publicado sensivelmente de 10 em 10 anos e constitui uma referência de grande qualidade no domínio da investigação em ensino, apresentando frequentemente «estados da arte» de várias áreas e domínios de investigação. Finalmente, refira-se que a Kluwer publicou, em dois volumes, o *International Handbook of Educational Evaluation* (Kellaghan e Stufflebeam, 2003), com quatro capítulos dedicados à avaliação das aprendizagens: um de Mislevy, Wilson, Ercikan e Chudowsky (2003), que discutem princípios psicométricos presentes na avaliação das aprendizagens dos alunos; um de Airasian e Abrams (2003), que apresentam e discutem cinco áreas de avaliação na sala de aula que se realizam antes, durante e depois do ensino e que, na opinião dos autores, representam a maioria das avaliações que os professores desenvolvem; um de Gipps e Stobart (2003), em que se enunciam e discutem as características de uma avaliação alternativa a partir do contraste entre os paradigmas psicométrico e educativo; e um de Kellaghan e Madaus (2003), em que se discutem exaustivamente as características, as formas de desenvolvimento, a validade, a fiabilidade, a comparabilidade, a equidade, as vantagens e as desvantagens dos exames nacionais mas também as vantagens e desvantagens das avaliações internas, baseadas nas escolas.

TRÊS RAZÕES SUFICIENTES PARA MUDAR A AVALIAÇÃO

Em praticamente todas as referências acima indicadas é clara a preocupação dos autores em demonstrar que é necessário mudar e melhorar as práticas de avaliação das aprendizagens dos alunos. Quer ao nível das salas de aula e das escolas quer ao nível da chamada avaliação em larga escala, como é o caso dos exames nacionais, obrigatórios, do ensino secundário ou das provas de aferição. E porquê? Julgo que essencialmente por três ordens de razões, todas igualmente importantes. (A ordem de apresentação é meramente arbitrária.)

Primeira Razão: Aprendizagem

Os resultados da investigação no campo das ciências cognitivas mostram que os processos de aprendizagem não são lineares, antes se desenvolvem em múltiplas direcções e a ritmos que não obedecem propriamente a padrões regulares. O desenvolvimento de processos complexos de pensamento deve iniciar-se logo desde o início da escolaridade e, em geral, não ocorre só após os alunos «dominarem» previamente um conjunto de factos básicos. As pessoas de todas as idades e de todos os níveis de competência utilizam, reconstroem e integram conceitos de diferentes graus de complexidade. Por outro lado, parece haver uma grande variedade nas formas e ritmos com que as pessoas aprendem, nas capacidades de atenção e de memória que podem utilizar nos seus diferentes desempenhos e na aprendizagem de conceitos e ainda nas formas que utilizam para comunicar os significados pessoais que atribuem ao que vão aprendendo. O que hoje sabemos acerca da aprendizagem permite-nos considerar inadequado, sob muitos pontos de vista, um ensino baseado quase exclusivamente na prática de procedimentos rotineiros e na aprendizagem de conhecimentos de factos discretos e descontextualizados que não são vistos de forma integrada. A investigação tem sugerido que aprender (?) desta forma dificulta a aplicação e mobilização dos conhecimentos em contextos diversificados, nomeadamente na resolução de problemas da vida real.

As aprendizagens significativas são reflexivas, construídas activamente pelos alunos e auto-reguladas. Por isso, os alunos não são encarados como meros receptores que se limitam a «gravar» informação, mas antes como sujeitos activos na construção das suas estruturas de conhecimento. Conhecer alguma coisa significa ter que a interpretar e ter que a relacionar com outros conhecimentos já adquiridos. Além disso, hoje reconhece-se que não basta saber como desempenhar uma dada tarefa mas é preciso saber quando a desempenhar e como adaptar esse desempenho a novas situações.

Sabe-se ainda que as competências metacognitivas e sócio-afectivas desempenham um papel relevante no desenvolvimento das aprendizagens. É por isso que, muitas vezes, as diferenças entre alunos fracos e bons na resolução de problemas não residem tanto nos conhecimentos que ambos os grupos possuem mas antes na utilização que ambos fazem desses mesmos conhecimentos. Ou seja, adquirir conhecimentos é uma condição necessária mas não é suficiente para que alguém se torne bom a resolver problemas. É preciso saber mobilizar atitudes e estratégias e saber quando e como as utilizar.

Parece evidente que, nestas circunstâncias, avaliar o desempenho dos alunos em tarefas rotineiras ou conhecimentos que pouco mais exigem do que a sua memorização será insuficiente. A avaliação tem que abranger processos complexos de pensamento, tem que contribuir para motivar os alunos para a resolução de problemas, tem que contribuir para que os aspectos de natureza sócio-afectiva sejam devidamente valorizados, tem que se centrar mais nas estratégias metacognitivas utilizadas e a utilizar pelos alunos.

Segunda Razão: Currículo

Nos últimos 30 anos, tem havido profundas alterações nos currículos de praticamente todos os países do mundo. A expansão dos regimes democráticos, a grande facilidade de mobilidade de pessoas e de mercadorias e as novas tecnologias de informação e de comunicação transformaram significativamente as sociedades modernas. Só a título de exemplo, pensemos nas profundas transformações económicas, sociais, políticas e culturais a que temos assistido nos países europeus. As sociedades tornaram-se muito mais multiculturais, muito mais competitivas, muito mais exigentes a todos os níveis e muito mais abertas do ponto de vista económico. Mas também se tornaram mais instáveis, mais incertas, mais inseguras, menos previsíveis. Hoje sabe-se que dificilmente um emprego será para toda a vida, numa dada empresa ou mesmo num dado país.

Há muito maior interdependência entre os países e, em particular, entre as suas economias. Decisões tomadas num país asiático podem ter reflexos políticos, económicos e sociais profundos nos países europeus. Podemos dizer que as mudanças rápidas e profundas, a imprevisibilidade e a interdependência económica, política e social são características marcantes das nossas sociedades. Os jovens de hoje têm que estar habilitados com um conjunto de saberes, capacidades e atitudes que lhes permitam viver integrados neste tipo de sociedades.

Nestas circunstâncias e numa tentativa de adaptar os sistemas educativos ao actual estado das coisas, mas também ao que se pensa que será o futuro, os governos têm procurado imprimir um ritmo reformista quase permanente que, essencialmente, se tem caracterizado pela introdução de alterações curriculares mais ou menos profundas, pela diversificação e flexibilização de percursos educativos e formativos, pela aposta na educação e formação ao longo da vida, pelo alargamento da escolaridade obrigatória, pela maior autonomia das escolas e, simultaneamente, pelo estabelecimento de mecanismos que, centralmente, reforcem o controlo do currículo e da avaliação por parte do Estado. A ideia parece ser a de procurar garantir que os sistemas educativos preparem jovens capazes de lidar com situações complexas, de resolver problemas em contextos diversificados nacionais e internacionais, de trabalhar em grupos cada vez mais multiculturais, de comunicar facilmente na sua língua e, pelo menos, numa língua estrangeira, de saber utilizar as novas tecnologias da informação, de seleccionar informação, de ser tolerantes para com as diferenças, de ser cidadãos críticos e participativos na vida democrática das sociedades. Enfim, a lista poderia prosseguir por mais algumas páginas... O que sobretudo interessa salientar é que os currículos de hoje nos lançam desafios que vão muito para além da memorização de conhecimentos e de procedimentos rotineiros. São muitíssimo mais exigentes do que há 30 anos atrás, quer na diversidade e profundidade de conhecimentos que se exigem quer na complexidade das tarefas que se propõem aos alunos quer ainda na preocupação explícita com a integração, relação e mobilização de conhecimentos e aprendizagens que, tanto quanto possível, se devem desenvolver em contextos com real significado para os alunos.

É neste contexto que os países participam em estudos internacionais de avaliação das aprendizagens dos alunos. Num certo sentido, tais estudos acabam por constituir uma pressão para que os países participantes adiram a certos padrões curriculares e para que os seus sistemas educativos definam *standards* que, supostamente, os ajudem a melhorar a qualidade do serviço prestado.

Curiosamente, só muito recentemente, mais concretamente a partir do início dos fins dos anos 80 e apenas num reduzido número de países, se começaram a perceber «movimentos», mais ou menos institucionais, tendentes a «alinhar» a avaliação com as novas exigências curriculares. Como veremos mais adiante, tal «alinhamento» ainda está longe de ser concretizado. Há um intenso e interessantíssimo debate que ainda está a fazer o seu caminho. Uma coisa parece ser certa: ninguém está propriamente satisfeito com o tipo de avaliação das aprendizagens dominante nos sistemas educativos. Há até quem diga, talvez exagerando um pouco, que as sucessivas reformas têm falhado porque a avaliação permanece praticamente imutável. Talvez não seja rigorosamente assim, mas a verdade é que, no caso do nosso país, só 18 anos após a instituição da democracia se pode dizer que surgiu um conjunto de orientações para a avaliação, baseado em princípios cognitivistas e construtivistas, mais consistente com uma visão progressiva da educação, do ensino e da aprendizagem. Trata-se do Despacho n.º 98-A/92.

Terceira Razão: Democracia

O acesso de todas as crianças e jovens ao bem da educação é uma conquista das sociedades democráticas. Percebe-se que, em última análise, a consolidação e o desenvolvimento das democracias depende do que formos capazes de fazer nos domínios da educação, da ciência e da cultura. Num certo sentido, tudo o resto poderá ser uma consequência do investimento que fizermos naqueles domínios da actividade humana. Acontece que ter todas, ou praticamente todas, as crianças e jovens nas escolas é uma importante conquista das sociedades, mas temos que pensar de que formas é que estão nas escolas. Será que todos os alunos têm as mesmas oportunidades para aprender? Será que todos recebem os mesmos tipos de *feedback* quanto aos seus progressos e dificuldades? Será que todos podem ir tão longe quanto as suas motivações, interesses e saberes lho permitirem? Será que todos, nas suas diferenças, sejam elas quais forem, se sentem plenamente integrados e vêem satisfeitas as suas legítimas aspirações? Será que a todos é proporcionada uma educação e uma formação que lhes permitam integrar-se plena e dignamente na sociedade? São questões que se colocam frequentemente e cujas respostas estão ainda longe de ser francamente positivas.

Mas o que é que a necessidade de mudarmos e melhorarmos a avaliação das aprendizagens dos alunos tem a ver com a democratização dos sistemas educativos? Eu diria que muito! Na verdade, as formas de organizar a avaliação podem motivar ou desmotivar os alunos, podem constituir importantes alavancas para superar obstáculos ou ser, elas mesmas, mais um obstáculo a superar, podem ajudar os alunos a estudar e a compreender bem as suas limitações e potencialidades ou, muito simplesmente, desinteressá-los. A avaliação pode e deve ter um papel relevante no desenvolvimento de aprendizagens complexas, no desenvolvimento moral e no desenvolvimento sócio-afectivo dos alunos. A avaliação pode segregar ou pode integrar. Pode melhorar a auto-estima dos alunos, pode piorá-la ou, em casos extremos, pode mesmo destruí-la. Pode orientar o percurso escolar dos alunos ou pode afastá-los de qualquer percurso!

Para muitos autores, a avaliação é uma das forças que mais influencia o desenvolvimento dos sistemas educativos porque:

1. Influencia as ideias dos estudantes acerca do que é importante aprender;
2. Afecta a sua motivação e a sua percepção acerca dos saberes, capacidades e atitudes a desenvolver;
3. Estrutura a forma como os alunos estudam e o tempo que dedicam ao trabalho académico;
4. Consolida as aprendizagens;
5. Promove o desenvolvimento dos processos de análise, síntese e dos processos meta-cognitivos.

Quaisquer mudanças e melhorias que se queiram introduzir nos sistemas educativos no sentido da sua integral democratização têm necessariamente que ser acompanhadas de esforços que nos permitam repensar a teoria e a prática da avaliação das aprendizagens. Só talvez dessa forma poderemos aspirar a que as escolas possam responder aos legítimos interesses e direitos das crianças, aos interesses e aspirações das comunidades em que vivem e, como acima se referiu, aos interesses das sociedades democráticas em que queremos viver.

CONCEPTUALIZAÇÕES DE AVALIAÇÃO

Guba e Lincoln (1989) distinguem quatro gerações de avaliação que, na sua opinião, correspondem a outras tantas perspectivas, abordagens, significados ou conceptualizações, que é possível identificar ao longo dos últimos cerca de cem anos. Os autores referem que a evolução dos significados que se foram atribuindo à avaliação não se pode desligar dos contextos históricos e sociais, dos propósitos que se pretendiam alcançar ou das convicções filosóficas dos que tinham algo a ver com a concepção, desenvolvimento e concretização das avaliações. Referem ainda que, ao longo dos tempos, as conceptualizações de avaliação se tornaram mais complexas e sofisticadas.

Muito sucintamente, caracterizam-se a seguir as três primeiras gerações identificadas e discutidas por Guba e Lincoln.

A Avaliação como Medida

Na primeira geração, conhecida como a geração da medida, avaliação e medida eram sinónimos. Isto é, a ideia que prevalecia era a de que a avaliação era uma questão essencialmente técnica que, através de testes bem construídos, permitia medir com rigor e isenção as aprendizagens escolares dos alunos. A inspiração para esta concepção vem dos testes destinados a medir a inteligência e as aptidões que se desenvolveram em França por Alfred Binet e que vieram a dar origem ao chamado coeficiente de inteligência, resultante do quociente entre a «idade mental» e a «idade cronológica» das pessoas. Este tipo de «testes mentais» foi largamente utilizado para fins de recrutamento, encaminhamento e orientação de jovens para as forças armadas e acabaram por se tornar cada vez mais populares nos sistemas educativos nos primórdios do século XX.

Refiram-se aqui dois factores que, na opinião dos autores, influenciaram esta primeira geração de avaliação. Um tem a ver com uma questão de afirmação dos estudos sociais e humanos que se começavam a realizar em Inglaterra, nos Estados Unidos, na Alemanha e em França, particularmente no contexto dos sistemas educativos e dos sistemas de saúde. Os marcantes e significativos sucessos da matemática e das ciências experimentais, que ocorreram ao longo do século XVIII e princípios do século XIX, foram também o sucesso dos métodos que lhes eram próprios, em particular o método científico. A investigação em ciências sociais, sem método e sem uma base sistemática de «ataque» aos problemas, era pressionada ou aconselhada, nomeadamente por Stuart Mill, a seguir o método experimental, no sentido de se afirmar junto da comunidade científica e de ganhar credibilidade (Guba e Lincoln, 1989; Madaus e Stufflebeam, 2000). Ora os testes e outros instrumentos destinados a medir aptidões ou aprendizagens humanas permitiam quantificá-las, compará-las ou ordená-las numa escala. De facto, era possível trabalhar matematicamente os seus resultados e proceder a um conjunto de transformações que poderiam servir uma variedade de finalidades. Esta quantificação das aprendizagens, das aptidões ou das inteligências dos alunos permitia seguir o modelo científico e obter a credibilidade que se pretendia para os estudos sociais e humanos.

O outro factor que acabou por ter uma importante influência no desenvolvimento e utilização dos testes para fins educativos foi a emergência do movimento da gestão científica no mundo da economia. O que, no fundo, se procurava era tornar o mais eficiente, eficaz e produtivo possível o trabalho dos seres humanos através de métodos de gestão que não cabe no âmbito deste texto especificar. A sistematização, a standardização e a eficiência caracterizavam o essencial deste movimento que teve em Fredrick Taylor o seu principal teórico. Madaus e Stufflebeam (2000), ao caracterizarem as diferentes «idades» da avaliação que, na sua opinião, se desenvolveram a partir de 1792 até aos nossos dias, consideram o período compreendido entre 1900 e 1930 como a *Idade da Eficiência e dos Testes* e sublinham a influência do trabalho e das ideias de Taylor no mundo da educação. Na verdade, as concepções essenciais do «Taylorismo» foram rapidamente adoptadas pelos sistemas educativos que, para muitos educadores e responsáveis, passaram a ser vistos como análogos às organizações empresariais. Ora os testes acabavam por ter um papel determinante para verificar, para medir, se os sistemas educativos «produziam» bons «produtos» a partir da matéria prima disponível – os alunos. Só nos Estados Unidos foram produzidos nesta altura milhares de testes standardizados. Uma bibliografia dedicada aos testes contava, em 1933, com mais de 3500. Em 1945, a autora, Gertrude Hil-dreth, publicou outra com 5200 entradas!

As concepções que são características desta geração ainda se mantêm e têm considerável influência nos sistemas educativos actuais. Trata-se, como vimos, de uma conceptualização em que avaliação e medida são sinónimos. Em termos práticos, de sala de aula, pode significar que a avaliação se reduz a pouco mais do que a administração de um ou mais testes e à atribuição de uma classificação em períodos determinados. Ou seja, uma perspectiva em que:

1. Prevalecem as funções sumativa, classificativa e selectiva da avaliação;
2. O único objecto de avaliação são os conhecimentos;
3. Há pouca, ou nenhuma, participação dos alunos no processo;
4. A avaliação é, em geral, descontextualizada;
5. Se privilegia a quantificação das aprendizagens em busca da objectividade e da neutralidade do professor (avaliador);
6. A avaliação é referida a uma norma ou padrão (por exemplo, a média) e, por isso, os resultados de cada aluno são comparados com os de outros grupos de alunos.

A Avaliação como Descrição

A segunda geração procurou superar algumas das limitações entretanto detectadas nas avaliações da primeira geração. Uma delas está directamente relacionada com o facto de os conhecimentos dos alunos serem considerados os únicos objectos de avaliação. Obviamente que, a certa altura, acabou por se considerar que era limitador avaliar um sistema educativo apenas com base nos resultados dos alunos. Há muitos outros intervenientes que têm que ser considerados e envolvidos num processo que, por exemplo, tenha a ver com a revisão dos currículos existentes. Guba e Lincoln (1989) referem que os avaliadores, perante objectivos educacionais previamente definidos, tinham como principal objectivo descrever padrões de pontos fortes e de pontos fracos. Por isso se referem à geração da descrição, que não se limita a medir, mas vai um pouco mais além ao descrever até que ponto os alunos atingem os objectivos definidos. A medida deixou de ser sinónimo de avaliação. Passou a ser um dos instrumentos ao seu serviço e, por isso, muitas das perspectivas anteriores mantêm-se presentes nas abordagens de avaliação. Ralph Tyler, um investigador e avaliador norte-americano, é referido como tendo tido uma grande influência nesta geração pois foi ele quem, pela primeira vez, se referiu à necessidade de se formularem objectivos para que se pudesse definir mais concretamente o que se estava a avaliar. Tyler, já nos anos 30 e 40, tinha uma concepção de currículo como um conjunto planeado e alargado de experiências formativas que ocorriam na escola, destinadas a contribuir para que os alunos pudessem atingir um conjunto de objectivos comportamentais previamente definidos. Avaliação educacional foi a expressão que escolheu para designar o processo de avaliar em que medida os objectivos eram ou não alcançados. Para muitos autores e investigadores (e.g., Guba e Lincoln, 1989; Madaus e Stufflebeam, 2000) Ralph Tyler é referido como o «pai da avaliação educacional» pelo trabalho que desenvolveu ainda nos anos 30 e 40 e que, ao longo de 25 anos, veio a ter uma influência muito significativa na educação e na avaliação, particularmente a partir da utilização de testes. Talvez por isso mesmo Madaus e Stufflebeam (2000) designam o período de avaliação decorrido entre 1930 e 1945 como a *Idade Tyleriana*. Parece evidente que a influência das perspectivas de Tyler é visível em muitas das avaliações que se desenvolvem nos sistemas educativos actuais. A grande diferença em relação à conceptualização anterior é o facto de se formularem objectivos comportamentais e de se verificar se eles são ou não atingidos pelos alunos. Pode talvez falar-se numa função reguladora da avaliação e na preocupação em conceptualizar o currículo de forma abrangente. Mas persistem todas as outras características da avaliação da geração anterior.

A Avaliação como Juízo do Valor

A terceira geração, designada por Guba e Lincoln (1989) como a geração da formulação de juízos ou julgamentos, nasce, tal como a anterior, da necessidade de superar falhas ou pontos fracos na avaliação da geração precedente. Sentiu-se que se deveriam fazer esforços para que as avaliações permitissem formular juízos de valor acerca do objecto de avaliação. Assim, os avaliadores, mantendo as funções técnicas e descritivas das gerações anteriores, passariam também a desempenhar o papel de juízes. Apesar dum conjunto de

reações quanto à questão da avaliação implicar a emissão de juízos de valor, a verdade é que, a partir dos finais da década de sessenta, todas as abordagens de avaliação, independentemente das suas diferenças, estavam de acordo nesse ponto. Esta geração de avaliação fica marcada pelo lançamento do *Sputnik*, pela então União Soviética, no ano de 1957. O Ocidente temia estar a ficar para trás na corrida ao espaço e receava que o desenvolvimento científico e tecnológico da União Soviética fosse muito superior. Houve então uma generalizada reação com uma importante expressão no desenvolvimento de reformas educativas orientadas para promover, em especial, o ensino da matemática e das ciências. Os países mais desenvolvidos da Europa e os Estados Unidos fizeram investimentos muito significativos na avaliação dos currículos, dos projectos e das aprendizagens dos alunos porque se queriam certificar de que os novos currículos obedeciam aos critérios de qualidade que se pretendiam alcançar. Foi uma época de grande expansão e desenvolvimento da avaliação e, talvez por isso, Madaus e Stufflebeam (2000) chamaram a este período, entre 1958 e 1972, a *Idade do Desenvolvimento*.

É, de facto, nesta geração que a avaliação alarga muito os seus horizontes. Surge, em 1967, através de Michael Scriven, a distinção entre o conceito de avaliação sumativa, mais associada à prestação de contas, à certificação e à selecção e o conceito de avaliação formativa, mais associada ao desenvolvimento, à melhoria das aprendizagens e à regulação dos processos de ensino e de aprendizagem (Nevo, 1986). Surgem as ideias de que a avaliação deve induzir e/ou facilitar a tomada de decisões, a recolha de informação deve ir para além dos resultados que os alunos obtêm nos testes, a avaliação tem que envolver os professores, os pais, os alunos e outros intervenientes, os contextos de ensino e de aprendizagem devem ser tidos em conta no processo de avaliação ou de que a definição de critérios é essencial para que se possa apreciar o mérito e o valor de um dado objecto de avaliação.

Síntese e Limitações das Três Gerações de Avaliação

Ao longo das três gerações que acima se discutiram sucintamente verificou-se que a avaliação se foi tornando mais complexa e mais sofisticada, evoluindo muito ao nível dos métodos utilizados, dos objectos de avaliação considerados ou dos propósitos ou finalidades. De uma concepção inicial muito redutora foi-se evoluindo para uma concepção mais sistémica e abrangente com a sistemática apreciação do mérito e do valor dos objectos avaliados, que deixaram de ser exclusivamente as «coisas» relativas aos alunos para passarem a incluir professores, projectos, currículos, programas, materiais, ensino ou políticas.

Em suma, poder-se-á dizer que, nesta altura, era relativamente consensual conceber a avaliação como um processo deliberado e sistemático de recolha de informação acerca de um ou mais objectos, no sentido de se poder formular um juízo acerca do seu mérito e/ou do valor que permitisse a tomada de decisões (*Joint Committee on Standards for Educational Evaluation*, 1981).

No entanto, segundo Guba e Lincoln (2000), todas as três gerações, no seu conjunto, apresentam três importantes limitações:

1. Uma tendência para as avaliações reflectirem os pontos de vista de quem as encomenda ou as financia, nunca, verdadeiramente, sendo postos em causa ou considerados co-responsáveis, mesmo que também tenham responsabilidades no sistema educativo. Normalmente, as responsabilidades pelos «falhanços» dos sistemas educativos tendem a ser distribuídas quase exclusivamente pelos professores e pelos alunos, ficando de fora todos os outros intervenientes;
2. Uma dificuldade das avaliações «acomodarem» a pluralidade de valores e de culturas existentes nas sociedades actuais;
3. Uma excessiva dependência do método científico ou, se quisermos, do paradigma positivista de investigação (Guba e Lincoln, 1994), que se traduz em avaliações pouco ou nada contextualizadas, com uma excessiva dependência da concepção de avaliação como medida e numa certa irredutibilidade das avaliações que se fazem porque estão apoiadas num método que, se bem utilizado, dá resultados muito dificilmente questionáveis. Do mesmo modo, os avaliadores nunca são postos em causa porque, considerando-se que a ciência é isenta de valores, a adesão ao método científico liberta o avaliador de quaisquer responsabilidades. Os avaliadores são neutros, não «contaminam» a avaliação nem se deixam «contaminar» por ela. Utilizam instrumentos que também são neutros e que medem com rigor o que são supostos medir.

Enfim, são algumas das perspectivas presentes na avaliação psicométrica, que também se caracteriza por estar associada a modelos de medida das aprendizagens, à standardização de tarefas de avaliação, de procedimentos de administração e de correcções e à utilização de instrumentos referidos à norma. Ou seja, o desempenho de um aluno é visto tendo em atenção o desempenho dos seus colegas e, nestas condições, é analisado em termos relativos e não em termos absolutos. Em última análise, dir-se-á que um estudante não pode controlar as suas próprias classificações porque não pode controlar os desempenhos dos seus colegas.

A Avaliação como Negociação e como Construção

Evidentemente que Guba e Lincoln (1989) propõem uma alternativa àquelas três gerações às quais atribuem as limitações que acabei de enunciar. Trata-se, naturalmente, da quarta geração de avaliação, que é uma geração de ruptura com as anteriores que, de acordo com aqueles autores, já não conseguem responder cabalmente às actuais necessidades de avaliação. Deve, no entanto, referir-se que os autores assumem que a sua própria concepção poderá ter dificuldades e limitações e que, eventualmente, no futuro, terá que ser revista nos seus pressupostos, concepções e métodos ou vir mesmo a ser completamente posta de parte. A quarta geração de avaliação caracteriza-se por ser *respondente*, isto é, à partida, não estabelece parâmetros ou enquadramentos. Estes serão determinados e definidos através de um processo negociado e interactivo com aqueles que, de algum modo, estão envolvidos na avaliação. Por outro lado, trata-se de uma avaliação construtivista, expressão que, no fundo, designa a metodologia que efectivamente é posta em prática na avaliação.

Noutros termos, poderemos dizer que a avaliação de quarta geração, *respondente* e construtivista, está baseada num conjunto de princípios, ideias e concepções de que se destacam os seguintes:

1. A avaliação é um conceito algo relativo que não tem propriamente uma definição, que é muito dependente de quem a faz e de quem nela participa;
2. Os professores devem partilhar o poder de avaliar com os alunos e outros intervenientes e devem utilizar uma variedade de estratégias, técnicas e instrumentos de avaliação;
3. A avaliação deve estar integrada no processo de ensino e aprendizagem;
4. A função formativa é a principal função da avaliação e está relacionada de perto com funções tais como a de melhorar, desenvolver, aprender ou motivar;
5. O *feedback*, nas suas mais variadas formas, frequências e distribuições, é um elemento indispensável na avaliação pois é através dele que ela entra no ciclo do ensino e da aprendizagem;
6. A avaliação deve servir mais para ajudar as pessoas a desenvolverem as suas aprendizagens do que para as julgar ou classificar numa escala;
7. A avaliação é um processo em que são tidos em conta os contextos, a negociação, o envolvimento dos participantes, a construção social do conhecimento e os processos sociais e culturais na sala de aula;
8. A avaliação deve utilizar métodos predominantemente qualitativos, não se pondo de parte a utilização de métodos quantitativos.

A avaliação de quarta geração, desenvolvida por Guba e Lincoln (1989), os trabalhos de Berlak (1992a; 1992b), de Broadfoot (1994), de Gipps (1994), de Gipps e Stobart (2003) e de tantos outros investigadores têm contribuído para que se vá afirmando a avaliação alternativa, também designada por avaliação autêntica, avaliação contextualizada ou avaliação educativa. Naturalmente que a avaliação alternativa surge baseada em concepções da aprendizagem diferentes das concepções subjacentes na avaliação psicométrica. Mas também surge baseada em concepções filosóficas diferentes ou, se preferirmos, paradigmas diferentes relativos à produção de conhecimento. Por isso me pareceu oportuno fazer aqui uma breve incursão nos paradigmas em que se baseiam as avaliações psicométricas e as avaliações alternativas.

AVALIAÇÃO PSICOMÉTRICA E AVALIAÇÃO ALTERNATIVA

Para Guba e Lincoln (1994), um paradigma é um sistema básico de concepções, de natureza axiomática, de que um indivíduo é possuidor e que lhe proporciona uma certa visão do mundo e da sua natureza. Acaba por ser, como refere Foucault (1979), um regime de verdade. Ou seja, um conjunto de práticas, procedimentos e discursos que uma dada comunidade científica utiliza e que acaba por determinar o que são, ou não, procedimentos legítimos de investigar ou de produzir conhecimento. O paradigma na sua função normativa dá indicações acerca dos métodos, procedimentos ou conhecimentos considerados, ou não, aceitáveis pela comunidade.

Para Guba e Lincoln (1994), as concepções básicas que definem um paradigma de investigação podem ser sintetizadas a partir das respostas que forem dadas a questões de natureza ontológica (Qual é a forma e a natureza da realidade? Há uma realidade susceptível de ser conhecida ou estudada de forma objectiva? O que é que existe que possa ser conhecido?), de natureza epistemológica (Qual a natureza da relação entre aquele que quer conhecer e aquilo que se pensa que pode ser conhecido? Como é que quem quer conhecer se relaciona com o conhecimento?) e de natureza metodológica (Que procedimentos são utilizados por quem quer conhecer para descobrir o que pensa que pode ser conhecido?).

Para Kuhn (1970), os paradigmas são sempre efémeros, são sempre transitórios. Os paradigmas transformam-se e evoluem através do desenvolvimento, por parte de uma dada comunidade científica, de novos conceitos, novas concepções e valores que se destinam a procurar responder a problemas e a questões a que o paradigma dominante já não responde de forma satisfatória. É então que surgem, no dizer de Kuhn, as «revoluções científicas» que arrastam a introdução de novas práticas e novas formas de abordar as questões e que se traduzem por rupturas mais ou menos dramáticas uma vez que o novo paradigma, ou o paradigma emergente, questiona e põe em causa todo o sistema de concepções, de valores e de interesses diversos que sustentam o velho paradigma.

Para muitos autores vive-se actualmente um momento de tensão motivado pelo facto do paradigma psicométrico, que é o que mais prevalece nos sistemas educativos e que está bem presente nas três primeiras gerações de avaliação definidas por Guba e Lincoln (1989), ter atingido um período crítico, pois não está a responder satisfatoriamente aos problemas de renovação e de melhoria do ensino, das aprendizagens e das escolas (e.g., Berlak, 1992a, 1992b; Berlak et al. 1992; Gipps, 1994; Gipps e Stobart, 2003). Estão de facto identificadas inconsistências entre o paradigma psicométrico e o que, por exemplo, hoje sabemos acerca das formas como os alunos aprendem, dos seus processos de pensamento e acerca dos aspectos sociais e culturais da cognição.

Vem a propósito referir aqui que vários autores têm vindo a sugerir aquilo a que alguns chamam a «inevitabilidade da convivência» do paradigma psicométrico, no âmbito da avaliação externa com efeitos na progressão dos alunos, com o paradigma da chamada avaliação alternativa, autêntica, educativa ou contextualizada, de inspiração construtivista e cognitivista, no âmbito da avaliação interna (Gipps, 1994; Gipps e Stobart, 2003; Madaus et al, 2000; Mislevy et al, 2003; Kellaghan e Madaus, 2000; Kellaghan e Madaus, 2003). Na realidade parece haver uma tendência cada vez maior para este «compromisso» nos sistemas educativos europeus. Pode haver várias explicações para esta situação. Desde as vantagens e desvantagens de cada uma das formas de avaliação quanto a questões de validade, de fiabilidade, de correcção, de equidade e de comparabilidade, até à evolução na elaboração de itens em exames ou provas de larga escala, ou à preocupação crescente em contextualizar os dados. A este propósito é interessante recordar aqui que, há 12 anos atrás, Berlak (1992a), Archbald e Newmann (1992) e Kilpatrick (1992) consideravam não ser claro assistir-se à rejeição do paradigma psicométrico e à sua substituição pelo paradigma da avaliação contextualizada. A tendência, diziam, poderá ir no sentido do desenvolvimento de uma teoria da avaliação baseada na evolução daqueles dois paradigmas. Algumas publicações recentes mostram que, em cerca de 10 anos, não se assistiu, de facto, à rejeição do paradigma psicométrico e que surgem variados exemplos de uma certa evolução no sentido de procurar uma conciliação com o paradigma construtivista tendo em conta as finalidades da respectiva avaliação. Ou seja, parece existir a tendência prevista por aqueles investigadores há cerca de 12 anos: a avaliação psicométrica e a avaliação alternativa (autêntica, educativa ou contextualizada) estão a evoluir e a «conviver» sob muitas formas num número aparentemente crescente de sistemas educativos (Kellaghan e Madaus, 2000, 2003; Kellaghan e Madaus, 2003; Kellaghan e Stufflebeam, 2003; Madaus et al, 2000; Mislevy et al, 2003; Torrance, 2003).

De acordo com Broadfoot (1994), uma das consequências da perspectiva psicométrica é a ênfase na avaliação de aspectos mais facilmente mensuráveis em detrimento de, por exemplo, processos mais complexos de pensamento, de atitudes e aspectos de natureza sócio-afectiva. Consequentemente, e tendo em conta, o que, por exemplo, nos dizem Berlak (1992a, 1992b), Lester e Kroll (1990), Raven (1992) e Schoenfeld (1985), quanto à indissociabilidade dos sistemas de concepções, valores, interesses e competências dos alunos e a sua influência nas aprendizagens, facilmente se infere que avaliar sem ter em conta estes sistemas pode ser insuficiente.

Por outro lado, aquilo a que muitos chamam a excessiva ênfase na utilização dos testes na avaliação psicométrica tem levado vários autores a identificar as limitações destes instrumentos. Na verdade, os testes em geral, apresentam, entre outras, as seguintes limitações:

1. Avaliam um leque relativamente estreito das competências previstas no currículo;
2. Tendem a fraccionar o conhecimento, assumindo a independência e não a interdependência dos objectivos educacionais;
3. Tendem a centrar-se em objectivos que suscitam mais a utilização de processos algorítmicos ou de procedimentos rotineiros e menos a utilização de processos complexos de pensamento, porque se pressupõe incorrectamente que os alunos devem, em primeiro lugar, atingir objectivos de nível elementar para só mais tarde atingirem objectivos de nível mais complexo;
4. Podem fornecer informação pouco válida e pouco fiável acerca do que se pretende avaliar (Broadfoot, 1994; Conceição, Neves, Campos, Fernandes, e Alaiz, sem data, Fernandes, 1991, 1992a, 1992b, 1993a, 1997; Lemos, Neves, Campos, Conceição e Alaiz, 1992; NCTM, 1995; Raven, 1992).

A consciência destas limitações dos testes, os emblemáticos instrumentos da avaliação psicométrica e os mais largamente valorizados e utilizados nas escolas, não deve fazer esquecer-nos as suas vantagens ou pontos fortes, tais como a de proporcionarem oportunidades para que os alunos mostrem saberes adquiridos, a de poderem constituir momentos de aprendizagem ou a de poderem ser elementos reguladores do processo de ensino-aprendizagem (Conceição et al., sem data).

É claro que, muitas vezes, os problemas poderão surgir mais das utilizações que se fazem dos testes e dos seus resultados do que dos testes em si mesmo. Os testes são necessários e podem ser instrumentos muito úteis no desenvolvimento das aprendizagens dos alunos. É necessário melhorar a natureza das questões, diversificar as possibilidades de administração, relativizar o seu peso no contexto de todos os elementos de avaliação que se recolhem e ter consciência de que, tal como todos os outros instrumentos, estratégias e técnicas de avaliação, os testes têm limitações.

Vejamos de seguida um conjunto de princípios orientadores para uma avaliação alternativa.

PRINCÍPIOS E ORIENTAÇÕES PARA UMA AVALIAÇÃO ALTERNATIVA

Perante a necessidade sentida de se desenvolver uma avaliação alternativa mais consentânea com as mudanças curriculares dos últimos anos, com as características sociais e culturais das escolas e ainda com os desenvolvimentos da psicologia das aprendizagens, importa caracterizar os seus princípios, os processos que utiliza, as formas que pode tomar e as respectivas utilizações.

Há um conjunto de princípios ou de orientações de natureza diversa que são mais ou menos referidos em toda a literatura sobre avaliação alternativa e que, de facto, contrastam com as limitações que vêm sendo apontadas à avaliação psicométrica. Estou a falar de princípios que têm sido adoptados em projectos de inovação (Anderson e Bachor, 1998; Archbald e Newmann, 1992; Bell, Burkhardt e Swan, 1992a, 1992b, 1992c; California Assessment Program, 1989; Koretz, 1998; Lange, 1987, 1993; Stecher, 1998; The Vermont Department of Education, 1992); recomendados em documentos programáticos (National Research Council, 1989, 1993; National Council of Teachers of Mathematics, 1989, 1991, 1995); ou enunciados em trabalhos de reflexão e/ou de investigação (Chambers, 1993; Cockroft, 1982; Conceição, 1993; Fernandes, 1991, 1992a, 1992b, 1993a, 1995; Gipps, 1994; Gipps e Stobart, 2003; Instituto de Inovação Educacional, 1992b; Lemos et al, 1992; Lester e Kroll, 1990; O' Connor, 1992; Raven, 1992; Romberg, Zarinnia e Collis, 1990; Tellez, 1996; Valencia, Hiebert e Afflerbach, 1994; Webb, 1992, 1993).

Selecionei apenas alguns que, naturalmente, me parecem importantes, embora tenha consciência de que outros ficam, nesta altura, por discutir, como é o caso da equidade que, de algum modo, é abordado noutra secção deste texto.

Integração Ensino-Aprendizagem-Avaliação

Uma nota claramente distintiva entre a avaliação psicométrica e a avaliação alternativa é o facto de, neste caso, o ensino, a aprendizagem e a avaliação constituírem como que um ciclo articulado e coerente. A avaliação, através do processo de comunicação que se estabelece e, muito particularmente, através de um *feedback* deliberada e devidamente preparado e utilizado, «entra» no ciclo do ensino e da aprendizagem. Como se costuma dizer é o *feedback* que contribui para a plena integração da avaliação, do ensino e da aprendizagem. Gipps e Stobart (2003) sustentam que a avaliação alternativa não fará qualquer sentido se não fizer parte integrante do processo ensino.

Uma adequada integração entre estes três processos permite, ou deve permitir, regular o ensino e a aprendizagem, utilizar tarefas que, simultaneamente, são para ensinar, aprender e avaliar e contextualizar a avaliação. Assim, haverá uma relação muito próxima entre as tarefas de avaliação e as finalidades do ensino. Como Gipps e Stobart (2003) sublinham, a avaliação alternativa não se limita exclusivamente à utilização de formas alternativas de avaliação mas é também uma utilização alternativa de avaliação como parte integrante do processo de aprendizagem.

Por outro lado, é através desta integração que se torna possível abranger mais consequentemente uma grande variedade de domínios do currículo ao contrário do que normalmente acontece com a avaliação psicométrica. Há mais consistência e mais equilíbrio entre as finalidades do currículo, o ensino e a avaliação.

Repare-se que, não havendo integração, a avaliação aparece como algo externo ao processo de ensino e de aprendizagem e como um procedimento cujas funções são de natureza mais sumativa, selectiva e certificadora e menos para ajudar os alunos a aprender e a desenvolver as suas aprendizagens.

Seleção de Tarefas

A selecção das tarefas é fundamental porque é através das actividades que suscitam quer aos professores quer aos alunos que pode residir a essência de um processo de ensino orientado para a resolução de problemas, para o desenvolvimento dos processos mais complexos de pensamento ou, numa palavra, para o desenvolvimento de um amplo espectro de aprendizagens que vão dos conhecimentos de conteúdos específicos da disciplina até aos aspectos de natureza mais transversal (e.g. sócio-afectivos, resolução de problemas, relação com os outros). É também através de tarefas adequadamente seleccionadas que se podem desenvolver processos de

avaliação mais contextualizados, mais elaborados, mais interactivos e mais directamente relacionados com a aprendizagem. De facto, as tarefas a utilizar podem facilitar a desejável articulação entre o ensino, a aprendizagem e a avaliação ou podem dificultar, ou mesmo impedir, essa articulação. Ou seja, sempre que possível, as tarefas devem ter uma tripla função:

1. Integrar as estratégias de ensino utilizadas pelo professor;
2. Ser meios privilegiados de aprendizagem;
3. Ter associado um qualquer processo de avaliação.

É preciso perceber-se que nem tudo tem a mesma importância. Ou seja, é preciso que se distinga o essencial do acessório, identificando os saberes, os conceitos e os procedimentos que são verdadeiramente estruturantes (aqueles que nenhum aluno deve deixar de trabalhar e de aprender, pois ajudam a estruturar áreas de saber e a estabelecer ligações com outras áreas) e, naturalmente, prever tarefas que reflectam aquela distinção. É preciso evitar, por todos os meios, a ideia de que tudo tem a mesma relevância e, ainda pior, que as pessoas se percam na definição de miríades de competências ou de objectivos e se deixem submergir numa atomização e fragmentação do currículo que, em tais circunstâncias, fará muito pouco sentido. É necessário que a avaliação reflecta as aprendizagens mais importantes, mais estruturantes e mais significativas a desenvolver pelos alunos. E isto significa, muito simplesmente, que é essencial a concentração de esforços na organização e desenvolvimento do ensino (e.g., selecção de tarefas, selecção das aprendizagens estruturantes, selecção de dinâmicas de sala de aula, selecção de estratégias de avaliação) a partir de um quadro tão simples quanto possível, o que não significa frágil, ligeiro ou simplista. Ocorrem-me, a propósito, palavras do consagrado arquitecto alemão Mies van der Rohe quando, com frequência, referia que *less is more*. Na verdade, parece-me importante trabalhar no sentido de integrar mais e dispersar menos. Dispersar pode ser um começo para não se fazer o essencial. Integrar obriga-nos a relacionar conceitos e domínios do currículo e a seleccionar o que é verdadeiramente importante. Por isso me parece oportuno dizer que, no fundo, não precisamos de mais avaliação. Precisamos de melhor avaliação! O que significa que talvez possamos evitar excessos, mais ou menos tecnicistas, nas salas de aula e fora delas, com a utilização de instrumentos que reflectem uma «pulverização» do currículo, como certas grelhas de observação, ou certas listas de verificação, que chegam a ter várias dezenas de itens para observar e para verificar! É talvez a melhor forma de não avaliarmos o que é verdadeiramente importante. A avaliação não é, de facto, uma mera questão técnica. É uma questão essencialmente pedagógica associada ao desenvolvimento pessoal, social e académico das pessoas. É uma questão de desenvolvimento curricular da maior relevância. É também uma questão ética. Por isso, antes dos instrumentos, que, naturalmente, interessa que sejam bem construídos e, sobretudo, melhor utilizados, vem o que pensamos sobre o papel que a avaliação deve desempenhar nas salas de aula, na educação e formação das crianças e dos jovens.

Na mesma linha de raciocínio, não precisamos de mais tarefas, precisamos é de melhores tarefas. Ricas do ponto de vista educativo e formativo, cuja resolução implique que os alunos relacionem, integrem e mobilizem um leque alargado de aprendizagens (e.g., conhecimentos, capacidades, atitudes, competências metacognitivas, competências sócio-afectivas). Sem prejuízo, obviamente, de se utilizarem uma variedade de tarefas que permitam alcançar os diferentes tipos de objectivos curriculares.

Entre as metodologias sugeridas para desenvolver o currículo inscrevem-se a necessidade de propor aos alunos tarefas de aprendizagem mais diversificadas e relacionadas com a vida real, a utilização de materiais manipulativos, o envolvimento em projectos destinados a resolver situações problemáticas ou o recurso ao trabalho de grupo. A concretização destas recomendações exige novas formas de avaliar. Uma prática de avaliação tradicional, quase exclusivamente baseada em testes de papel e lápis, seria insuficiente e até totalmente desajustada, pois tenderia a ignorar todas as competências que vão para além da aquisição de conhecimentos.

Por isso, deve haver consistência entre a avaliação, o currículo e as metodologias e estratégias utilizadas para o desenvolver, fazendo, deste modo, sempre que possível, coincidir as tarefas de aprendizagem com as tarefas de avaliação. Só desta forma, como vimos acima, a avaliação fará parte integrante do ensino e da aprendizagem e poderá assumir o seu papel regulador.

Funções

As funções da avaliação mais tradicionalmente referidas são as funções formativa e sumativa. A avaliação de natureza psicométrica privilegia a função sumativa, isto é, a formulação de juízos acerca das aprendizagens dos alunos **após** um período mais ou menos longo de ensino e de actividades conducentes à aprendizagem. É uma perspectiva consistente com uma concepção de avaliação como medida de produtos do ensino, através de instrumentos, normalmente testes, referidos à norma. Uma concepção que, consequentemente, tende a dar mais relevância às funções de classificação, de ordenação, de selecção ou de certificação e, por isso mesmo, não dá destaque aos processos de aprendizagem nem aos contextos em que esta se desenvolve. O que não significa que a avaliação psicométrica não considere de todo a avaliação formativa ou que não lhe dê qualquer atenção.

Por sua vez, a avaliação alternativa tende a dar mais destaque à avaliação formativa, isto é, aquela que tem lugar **durante** os períodos em que ocorrem o ensino e todas as actividades a ele associado. É uma concepção substancialmente diferente da anterior já que há uma preocupação com os processos de ensino e de aprendizagem, que são analisados *in loco*; com a plena integração da avaliação nesses mesmos processos; com a criteriosa selecção de tarefas que motivem e mobilizem os alunos; e com o seu envolvimento, tão activo quanto possível, na aprendizagem e na sua avaliação. Por isso, a avaliação alternativa tende a dar relevância a funções da avaliação tais como a motivação, a regulação e a auto-regulação, o apoio à aprendizagem, a orientação ou o diagnóstico.

É evidente que em qualquer processo de avaliação têm que se considerar os processos e os produtos de aprendizagem. Só que as ênfases, nas duas concepções aqui revistas, são diferentes e, como tal, deverão produzir resultados diferentes! A prática consequente da avaliação alternativa não implica que a certificação se faça e, implicitamente, a selecção. Mas sempre convém referir que a atribuição de classificações não é a única função da avaliação. É muito importante que a avaliação ajude a motivar os alunos para aprenderem e para lhes dar conta dos seus progressos e dos seus sucessos mas também dos seus insucessos e dificuldades.

Triangulação

Como facilmente se compreende não é possível avaliar tudo o que um aluno sabe e é capaz de fazer. O que normalmente fazemos é avaliar amostras de desempenhos dos alunos, relativamente a domínios previstos no currículo, na sequência da resolução de tarefas de natureza diversa. Em geral, se tudo correr bem, as amostras de desempenho numa variedade de tarefas permitem-nos afirmar com alguma segurança se os alunos aprenderam, ou não, o que era suposto aprender relativamente a um dado domínio.

Sabemos, por outro lado, que, sempre que avaliamos estamos a cometer um erro. Ou seja, não há nenhuma estratégia, técnica ou instrumento que nos permita avaliar exactamente determinadas aprendizagens dos alunos. Todos têm as suas vantagens e desvantagens e todos nos induzem num erro que, naturalmente, temos que procurar reduzir à sua ínfima expressão.

Também não é fácil garantir que a avaliação abranja todos os domínios do currículo ou mesmo o essencial de cada um dos domínios.

O desenvolvimento das aprendizagens dos alunos está condicionado por um conjunto complexo e interdependente de factores tais como as suas capacidades intelectuais, os seus sistemas de concepções, as suas capacidades metacognitivas, as suas atitudes, desejos, persistência ou os contextos sócio-culturais em que se inserem. Convém ter em atenção que os alunos possuem conhecimentos, aptidões, motivações, estilos e ritmos de aprendizagem que podem variar significativamente. Há alunos que resolvem situações problemáticas por escrito com facilidade mas podem não ter a mesma facilidade se têm que as resolver oralmente, ou perante o grupo da sua turma. Há alunos com uma inteligência linguística de fraco nível mas que podem ter uma inteligência lógico-matemática ou uma inteligência espacial de nível bom ou mesmo muito bom (Gardner, 1983; 1991; Gardner e Hatch, 1989).

A recolha de informação, que toda a avaliação pressupõe, tem que ter em conta as considerações que acima se fizeram. Por isso se propõe um princípio de triangulação aplicável às estratégias, técnicas e instrumentos, aos intervenientes no processo de avaliação, aos tempos ou momentos de avaliação e aos espaços ou contextos.

Triangulação de estratégias, técnicas e instrumentos

É necessário diversificar os métodos e instrumentos de recolha de dados e encontrar formas de dar alguma estrutura à avaliação de natureza mais informal. É através da avaliação que os alunos tomam consciência do tipo de actividades, experiências de aprendizagem, atitudes, valores, conhecimentos e competências que são valorizados. A utilização privilegiada de testes de papel e lápis é manifestamente insuficiente. É desejável que se recolha informação através de relatórios, de pequenos comentários, de observações mais ou menos estruturadas, de conversas (entrevistas) mais ou menos formais, mais ou menos estruturadas ou de trabalhos e produtos realizados pelos alunos de diversa natureza. Perante a diversidade de tarefas de avaliação os alunos percebem que não chega «estudar para o teste» e que se espera que desenvolvam um alargado leque de aprendizagens.

A diversificação de «métodos» de recolha de informação permite avaliar mais domínios do currículo, lidar melhor com a grande diversidade de alunos que hoje estão nas salas de aula e também reduzir os erros inerentes à avaliação.

Triangulação de intervenientes

A diversificação dos intervenientes no processo de avaliação justifica-se pelas razões já acima enunciadas mas também por razões tais como:

1. A tendência para se avaliar o trabalho do professor e da escola;
2. A necessidade de proceder a apreciações mais globais e transversais dos progressos experimentados pelos alunos;
3. A reconhecida importância da avaliação diagnóstica e formativa;
4. A necessidade de resolver os problemas de aprendizagem por meios pedagógicos, em vez de por meios administrativos.

Para poder dar resposta a tudo o que hoje dele se exige, o processo de avaliação deve poder contar com os próprios alunos, com os pais, com outros professores e, se necessário, com outros técnicos, como é o caso dos assistentes sociais e dos psicólogos escolares. Assegura-se deste modo a triangulação entre os principais intervenientes no processo de avaliação o que, como se imagina, tem consequências a diversos níveis. Um deles é o da disponibilidade dos professores para partilharem genuinamente o poder que lhes é conferido pela avaliação.

Triangulação de espaços e de tempos

Na mesma linha dos raciocínios anteriores, parece importante que a avaliação possa decorrer em diferentes contextos ao longo de diferentes períodos de tempo. Isto significa pelo menos duas coisas. Uma é a de que pode ser importante recolher informação dentro da sala de aula, onde se podem criar situações muito diversificadas (e.g., trabalho individual, em pequeno grupo, em grande grupo, inter-pares, dramatizações, simulações de conferências, apresentações, leitura, etc.) mas também fora dela como é, por exemplo, o caso de visitas a museus, a unidades empresariais, a mercados, a instituições científicas ou a instituições públicas. Outra é a de que a informação, desejavelmente, deve ser recolhida em tempos diversificados; sempre que possível ao longo dos períodos escolares e não em dois ou três momentos previamente anunciados.

Transparência

Qualquer processo de avaliação tem que ser transparente. Os objectivos, as aprendizagens a desenvolver e todos os processos de avaliação devem ser claramente expressos e devem estar sempre disponíveis para quem a eles quiser ter acesso.

Os critérios de avaliação devem ser apresentados de forma clara e devem constituir um elemento fundamental de orientação dos alunos. Ou seja, qualquer apreciação que se faça do trabalho dos alunos deve ter em conta os respectivos critérios e os alunos devem ser capazes de perceber sem problemas a sua situação face às aprendizagens que têm que adquirir ou desenvolver.

A ESSENCIALIDADE DO *FEEDBACK* NA AVALIAÇÃO ALTERNATIVA

Na avaliação alternativa, a comunicação, sob as mais diversas formas, assume um papel indispensável. Sem ela estaríamos num processo, mesmo assim pouco habitual, de avaliação psicométrica. É através da comunicação que todos os alunos devem tomar consciência dos seus progressos e/ou dificuldades em relação às aprendizagens que têm que adquirir. É através da comunicação que os professores também poderão perceber as alterações que necessitam de fazer para que o seu ensino vá ao encontro das necessidades dos seus alunos.

É evidente que, na ausência de qualquer informação por parte dos professores, os alunos terão certamente mais dificuldades em situar-se perante os desafios que têm que enfrentar, os esforços que têm que fazer ou os métodos de estudo que devem utilizar. Devolver qualquer trabalho a um aluno em branco, com uma mera designação ou com uma apreciação escrita são certamente situações muito diferentes. Os alunos precisam de orientações sistemáticas e de avaliações do seu trabalho e dos seus desempenhos que os ajudem a melhorar as suas aprendizagens, que os estimulem e que os motivem a ir tão longe quanto possível ou que reconheçam os seus progressos e sucessos. Ou seja, os alunos precisam de *feedback* acerca dos processos e produtos do seu trabalho e acerca dos seus comportamentos sociais.

Como já se referiu, o *feedback* é indispensável para que a avaliação integre os processos de ensino e de aprendizagem e, muito particularmente, para que a avaliação assuma a sua natureza formativa. De facto, através de um *feedback* regular e sistematicamente providenciado, os alunos podem começar a desenvolver competências de auto-avaliação e de auto-regulação das suas aprendizagens durante, e não no final, de um dado período de ensino e aprendizagem. Consequentemente, podem utilizar o *feedback* como orientação para melhorar ou corrigir o caminho que vinham seguindo.

O *feedback* desempenha um papel crucial na aprendizagem porque, através dele, os alunos são sistematicamente lembrados dos níveis de aprendizagem, ou dos *standards*, que é necessário alcançar e ficam cientes dos seus próprios progressos tendo em conta a comparação com os seus próprios desempenhos anteriores ou critérios previamente definidos.

Um dos aspectos que merece ponderação relaciona-se com a distribuição do *feedback*, seja ele de natureza escrita, oral ou não verbal. Pode haver a tendência para que, numa turma, o *feedback* seja mal distribuído pelos diferentes alunos. Pode acontecer que os bons alunos, os alunos que apresentam mais respostas certas, os rapazes, os alunos das etnias dominantes ou os alunos das classes média e média alta recebam mais *feedback* do que os alunos mais fracos, do que os que apresentam mais respostas erradas, do que as raparigas, do que os alunos de minorias étnicas ou do que os alunos das classes sociais mais desfavorecidas.

Por outro lado, para além de ser necessário distribuir equitativamente o *feedback*, é importante equilibrar a forma e a natureza que ele pode assumir. Convém equilibrar a proporção de *feedback* oral, escrito ou não verbal, a oportunidade para o fazer em público ou em privado, dentro ou fora da sala de aula ou de o focar exclusivamente nos aspectos referentes a conteúdos disciplinares específicos ou também noutros aspectos como, por exemplo, nas aprendizagens de natureza transversal.

Tunstall e Gipps (1996) distinguem entre *feedback* avaliativo, em que são emitidos juízos acerca do valor ou do mérito de um dado trabalho, com referência implícita ou explícita a normas ou *standards*, e *feedback* descritivo, relativo às tarefas apresentadas aos alunos e utilizado para se referirem explicitamente às aprendizagens evidenciadas na sua resolução. Os autores consideram dois tipos de *feedback* descritivo. Um, em que o professor se mantém essencialmente no controlo de toda a actividade pedagógica, nomeadamente mantendo todo o poder na avaliação, e em que se reconhecem as aprendizagens conseguidas pelos alunos, se produzem diagnósticos com base em critérios específicos ou se corrigem procedimentos. Outro, em que o professor partilha o seu poder de avaliador, com os alunos, responsabilizando-os igualmente pelas suas aprendizagens, analisando em conjunto a utilização de estratégias para a auto-regulação e auto-controlo do processo de aprendizagem. Por isso, este último tipo de *feedback* descritivo pode contribuir para que os alunos, sempre apoiados pelo professor, se tornem mais autónomos para avaliarem e regularem os seus desempenhos e para encontrarem maneiras de os melhorar. O professor acaba por assumir um papel mais «facilitador» em vez de mais «fornecedor» de *feedback* ou de «juiz». É, segundo Gipps e Stobart (2003), uma perspectiva em que se dá ênfase ao papel do professor *com* o aluno em vez de *para* o aluno.

A avaliação alternativa não o será verdadeiramente sem a utilização deliberada, sistemática e fundamentalmente didáctica e pedagógica, de um «sistema» de *feedback* que apoie, regule e melhore os processos de aprendizagem e de ensino. Desta forma, consolida-se a função formativa da avaliação, contribui-se para que os alunos se tornem mais autónomos, mais responsáveis pelas suas aprendizagens, mais capazes de avaliarem e regularem o seu trabalho, o seu desempenho e as suas aprendizagens e mais ágeis na utilização das suas competências metacognitivas.

O *PORTFOLIO* COMO EXEMPLO DE AVALIAÇÃO ALTERNATIVA

Os *portfolios* de trabalhos dos alunos têm sido referidos como uma estratégia que pode permitir a organização da avaliação alternativa de acordo com as ideias e os princípios que acima se apresentaram e discutiram.

Um *portfolio* é uma colecção organizada e devidamente planeada de trabalhos produzidos por uma aluna ou um aluno durante um certo período de tempo. A sua organização deve ser tal que permita uma visão tão alargada, tão detalhada e tão profunda quanto possível das aprendizagens conseguidas pelos alunos.

Não há propriamente um modelo ou um formulário para a construção de um *portfolio*. Professores e alunos, de acordo com os objectivos que se pretendem alcançar, as condições e os recursos de que dispõem, estabelecem as regras; isto é, acertam o tipo de trabalhos a incluir no *portfolio*, as condições em que o podem fazer e os objectivos de tal inclusão. O processo de avaliação deve também ser objecto de discussão e de negociação.

Os produtos elaborados pelos alunos, que devem constituir uma amostra alargada do que sabem e são capazes de fazer, podem ser de natureza muito diversificada e, naturalmente, devem contemplar os domínios destacados pelo currículo e/ou pelo projecto de escola e/ou pelo projecto curricular de turma. É por isso desejável que os trabalhos a integrar no *portfolio* tenham, pelo menos, as seguintes características:

1. Contemplem todos os domínios do currículo ou, pelo menos, os que são considerados essenciais e estruturantes;
2. Sejam suficientemente diversificados quanto à forma (escritos, visuais, audiovisuais, multimédia);
3. Evidenciem processos e produtos de aprendizagem;
4. Exemplifiquem uma variedade de modos e processos de trabalho;
5. Revelem o envolvimento dos alunos no processo de revisão, análise e selecção de trabalhos.

Um *portfolio* poderá incluir uma ampla variedade de trabalhos dos alunos tais como relatórios, composições, comentários breves a textos ou a quaisquer situações de aprendizagem (filme educativo, peça de teatro, acontecimento científico, tecnológico ou social), relatos e reacções escritas a visitas de estudo, trabalhos individuais e de grupo, produtos multimédia, desenhos e ilustrações, reflexões dos alunos sobre a disciplina, os seus progressos, as suas dificuldades, resoluções de problemas ou relatos de experiências. (É apenas uma lista de sugestões. As circunstâncias e os contextos em que as pessoas funcionam é que devem determinar o tipo de «coisas» a incluir.)

No fundo, a ideia de organizar, de forma pensada e deliberada, os trabalhos dos alunos num *portfolio*, é a de obter uma imagem, tão nítida quanto possível, das aprendizagens que desenvolveram ao longo de um dado período de tempo, das suas experiências, dificuldades, e progressos.

Um ambiente de ensino e de aprendizagem em que, frequentemente, os alunos têm uma participação activa, em que se propõe um conjunto diversificado de tarefas, em que se resolvem situações problemáticas ou em que se utilizam diferentes dinâmicas de trabalho exige uma avaliação diferente. Uma avaliação mais contextualizada ou mais autêntica, mais participada e mais reflexiva.

Mais contextualizada, ou mais autêntica, porque ocorre à medida que os alunos vão resolvendo ou trabalhando nas tarefas que lhes são propostas no contexto «normal» da sala de aula, sem tantos constrangimentos de tempo e de administração que têm que ser utilizados nos testes standardizados. Mas também porque os alunos têm oportunidades para analisarem o seu trabalho e, por isso, a auto-avaliação e auto-regulação das aprendizagens podem ser práticas habituais. Nestas condições, os alunos poderão, em princípio, ter mais oportunidades para mostrar o que sabem e são capazes de fazer e os professores mais oportunidades para conhecer as suas dificuldades e ajudá-los a superá-las.

Mais participada porque os professores partilham o poder da avaliação com os seus alunos, com outros professores e até com os pais e encarregados de educação. A avaliação é assim um meio para também envolver outros intervenientes no apoio aos alunos, ajudando-os a superar dificuldades, a delinear estratégias de estudo e de trabalho ou a apreciar os seus trabalhos. A participação dos alunos na avaliação pode ser um processo de grande valor educativo e formativo, pois contribui para que desenvolvam um importante conjunto de aprendizagens de natureza cognitiva e metacognitiva e de natureza social, cultural e afectiva.

Mais reflexiva pois há oportunidades para que os alunos se habituem a rever os seus trabalhos de forma crítica, consciente e sistemática. Consequentemente, podem analisar o que fizeram, identificar o que de mais característico existe no seu trabalho, o que foi evoluindo com o tempo ou o que ainda necessita de ser feito. Trata-se de desenvolver competências que lhes permitam rever e reformular o seu trabalho.

Uma avaliação com estas características pode proporcionar mais e melhor informação aos professores e responsabilizar mais os alunos pela sua própria aprendizagem, uma vez que são chamados a criticar e a reflectir, sistematicamente, sobre os seus próprios trabalhos.

Sublinhe-se bem que um *portfolio* não é uma mera colecção de trabalhos dos alunos «organizados» numa pasta de arquivo ou numa caixa. A organização do *portfolio* exige uma planificação com propósitos bem claros e uma articulação sistemática entre o desenvolvimento do currículo, a aprendizagem e a avaliação. Aderir a esta abordagem pode trazer vantagens tais como:

1. Abranger mais processos e objectos de avaliação;
2. Fazer coincidir tarefas de avaliação e de aprendizagem, contribuindo para melhorar a consistência entre o currículo, o seu desenvolvimento e a avaliação;
3. Contextualizar a avaliação na medida em que ela surge associada à situação em que a aprendizagem teve lugar e não numa situação mais formal, desligada do ambiente em que se trabalham as tarefas;
4. Mostrar mais acerca do que os alunos sabem e são capazes de saber;
5. Incentivar os alunos a participarem mais activamente no processo de avaliação e a reflectir criticamente sobre o seu próprio trabalho;
6. Melhorar a auto-estima dos alunos porque têm mais oportunidades para mostrar o que conseguem fazer e para evoluir;
7. Identificar mais facilmente os progressos e as dificuldades dos alunos dada a natureza longitudinal do *portfolio*;
8. Conhecer mais detalhada e profundamente as aprendizagens dos alunos o que permite tomar melhores decisões.

A utilização desta estratégia não é simples. Implica, como já se disse, uma planificação e organização rigorosas, uma revisão sistemática e regular dos trabalhos dos alunos e um cuidado muito especial com o tipo de tarefas que lhes queremos propor. Não há qualquer garantia de que a utilização de *portfolios* implique, por si só, uma avaliação alternativa, mais autêntica, mais participada e mais reflexiva. Na verdade, os *portfolios* podem facilmente tornar-se meras pastas com colecções de trabalhos dos alunos. No entanto, se bem utilizados, podem influenciar positivamente as formas como se ensina, se aprende e se avalia. Podem dar origem a uma outra cultura, a uma outra ideia, de sala de aula: um espaço em que as aprendizagens se vão construindo em conjunto e/ou individualmente ao ritmo de cada um, em que se reflecte e se pensa, em que se valorizam as experiências, intuições e saberes de cada aluno, em que se acredita que as dificuldades podem ser superadas e em que, essencialmente, se ensina e se aprende. Com mais ou menos esforço, mas sempre com gosto.

OS PROFESSORES E A AVALIAÇÃO DAS APRENDIZAGENS

As considerações que se seguem são baseadas numa análise de um conjunto de investigações realizadas com professores dos três ciclos da educação básica e do ensino secundário (Alves, 1997; Antunes, 1995; Boavida, 1996; Campos, 1996; Fernandes, Neves, Campos e Lalande, 1996; Gil, 1997; Lobo, 1996; Neves, 1996). Na maioria destas investigações, estudos de caso de natureza qualitativa, participou um número relativamente reduzido de professores, que foram entrevistados, e cujas aulas foram, em muitos casos, observadas. Naturalmente que o que aqui se pretende é, tanto quanto possível, integrar conclusões e reflexões constantes naquelas investigações numa tentativa de ilustrar como é que os professores participantes organizavam as suas práticas de avaliação das aprendizagens e o que é que pensavam sobre o assunto. Não há aqui qualquer ideia de apresentar esta integração de resultados como uma generalização a todos os professores. Pode no entanto referir-se que os resultados, assim como as interpretações consequentes, são, genericamente, consistentes com outras investigações e/ou reflexões de índole teórica (e.g., Archbald e Newmann, 1992; Berlak, 1992a; Berlak, 1992; Broadfoot, 1994; Conceição, 1993; Fernandes, 1997; Gipps, 1994; Gipps e Stobart, 2003; O'Connor, 1992; Raven, 1992; Stiggins e Conklin, 1992).

Em geral, todos os participantes nas investigações dão uma clara ênfase à avaliação de conhecimentos específicos das disciplinas que leccionam. Em particular, aos factos, conceitos e procedimentos constantes nos livros de texto. Competências de natureza transversal, como, por exemplo, as do domínio da resolução de problemas ou do domínio das relações sócio-afectivas, são raramente valorizadas e, consequentemente, raramente avaliadas. Há várias explicações possíveis para esta constatação:

1. A comunicação que se estabelece na sala de aula é essencialmente organizada em torno de conteúdos específicos que acabam por constituir, quase naturalmente, os principais, ou mesmo únicos, objectos de avaliação;
2. Para os professores participantes, as competências de natureza transversal, ao contrário dos conteúdos disciplinares, não estão clara e expressamente definidas nos currículos como alvos a atingir, por isso sentem falta de «objecto» de avaliação;
3. Culturalmente os professores tendem a valorizar mais, e quase exclusivamente, os conhecimentos específicos que os alunos evidenciam. De facto, não se conhecem casos de alunos que vejam posta em causa a sua progressão académica quando demonstram possuir conhecimentos considerados suficientes, apesar de eventualmente poderem revelar ausência de competências transversais como as de natureza sócio-afectiva ou outras.

Os testes são claramente os instrumentos mais valorizados e utilizados pelos professores participantes, que tendem a avaliar exclusivamente conhecimentos correspondentes aos conteúdos de natureza académica constantes nos programas. Penso tratar-se de um resultado que faz sentido e que é coerente com as concepções sobre o ensino, a aprendizagem e a avaliação sustentadas por estes professores, que acabam por sentir que os testes lhes permitem medir com rigor as aprendizagens dos alunos. Sentem-se mais seguros e também sentem que a sua vida como avaliadores fica menos complicada. No entanto, estes professores acabam por recolher uma quantidade significativa de informação sobre o desempenho e sobre os saberes dos seus alunos que, muito provavelmente, é utilizada de forma algo aleatória, sem se inscrever numa estratégia coerente e deliberada de avaliação. Trata-se de informação que é essencialmente proveniente de observações e de conversas informais. Por outro lado, os professores que, deliberadamente, também avaliam outros saberes e outras capacidades tendem a ver os testes como «mais um meio para recolher informação e evidências de aprendizagem» e não «o meio privilegiado para avaliar».

De modo geral, verifica-se que a atribuição de classificações está sobretudo dependente dos resultados dos testes. Os professores revelaram ter dificuldades várias na integração da informação que têm disponível, que é de natureza mais ou menos diversificada. Muito provavelmente porque não lhes é igualmente fácil estabelecer critérios que os ajudem a fazer corresponder as aprendizagens dos alunos a um dado valor de uma dada escala. Nestas condições, os professores acabam por encontrar procedimentos ou «fórmulas» mais ou menos consistentes, resultantes da sua intuição, conhecimento e experiência. Em todo o caso, os resultados dos testes, percebidos como mais objectivos, mais válidos e mais fiáveis, acabam por ter um peso eventualmente desproporcionado nas decisões que precedem a atribuição de classificações finais.

A grande maioria dos participantes nas investigações referidas concorda com a necessidade de se diversificarem as estratégias, técnicas e instrumentos de avaliação. Verificou-se, no entanto, que as práticas

de muitos professores não eram consonantes com aquela necessidade. É interessante assinalar que os professores em cujas aulas se desenvolvia um ambiente em que os alunos participavam mais activamente na aprendizagem pareciam ter mais facilidade em valorizar e em utilizar outras estratégias, técnicas e instrumentos para além dos testes (e.g., *portfolios*, observações sistemáticas mais ou menos estruturadas, pequenos relatórios, trabalhos de casa, questionários).

A ideia com que se fica é a de que a avaliação é mais diversificada, mais significativa do ponto de vista educativo e formativo, mais consistente com o ensino e com a aprendizagem, quando os professores organizam o ensino e promovem um ambiente de trabalho em que os alunos participam mais activamente na resolução de uma diversidade de tarefas cuidadosamente seleccionadas. A avaliação surge então como um processo que ajuda a melhorar, a corrigir, a integrar, a regular, a definir critérios, a diagnosticar e a facilitar a comunicação. Um processo que é indissociável do ensino e da aprendizagem. Deve no entanto referir-se que esta maneira de trabalhar, isto é, de ensinar e de avaliar, parece resultar mais da intuição, da experiência e dos saberes que dela são decorrentes e menos de um processo deliberadamente consciente, ancorado em perspectivas teóricas que se procuram pôr em prática.

Outro resultado interessante é o de que a maioria dos professores participantes nas investigações não explicitam os critérios de avaliação junto dos seus alunos. Mesmo os que o fazem actuam de forma mais ou menos isolada. Ou seja, os critérios são definidos um pouco de acordo com o que cada um pensa acerca do que é importante e não parecem resultar de uma «política» definida ao nível dos órgãos próprios da escola (e.g., conselho pedagógico, departamento curricular). Por razões várias é fundamental que os critérios de avaliação sejam partilhados e até discutidos com os alunos. A avaliação tem que ser transparente e os critérios ajudam os alunos a organizar o seu estudo, contribuem para os motivar a aprender e a delinear estratégias de aprendizagem e de envolvimento nas tarefas que lhes são propostas pelos professores.

As investigações analisadas permitem que se fique com a ideia de que, em geral, os professores pensam, vivem e praticam a avaliação de forma muito isolada; este facto pode ajudar a explicar a relativa falta de confiança que os professores manifestam relativamente às suas avaliações e às decisões que delas decorrem assim como aos procedimentos que utilizam. Uma «política» de avaliação de escola que integre as principais preocupações dos professores e que valorize as suas experiências e saberes pode ser uma ideia com bom acolhimento. O debate e a reflexão entre os professores de uma escola, de uma turma ou de um grupo de turmas poderia contribuir para clarificar áreas problemáticas tais como a definição de critérios de avaliação, a diversificação de estratégias, técnicas e instrumentos, a integração de informação, a atribuição de classificações, a identificação de funções de avaliação, o envolvimento dos alunos, dos pais e de outros intervenientes, a distribuição, a frequência e a natureza do *feedback* ou as questões de validade, fiabilidade e de equidade. Parece claro que os professores e as próprias escolas precisam de estar menos sós com as suas avaliações.

Em síntese, a ausência de «políticas» de ensino e de avaliação nos projectos educativos das escolas e/ou nos projectos curriculares de turma parece ser uma importante debilidade que merecia ser cuidadosamente enfrentada. Os professores parecem estar um pouco entregues a si próprios numa questão que tem uma grande relevância pedagógica, didáctica e educativa mas tem também uma grande relevância social. (É bom que se comecem a investigar e a discutir os diferentes tipos de consequências pessoais e sociais resultantes das avaliações.) Talvez por isso, mesmo os professores sentem-se relativamente inseguros e acabam por fazer as «coisas» sem possuírem uma visão ampla e clara acerca do que realmente poderá estar em causa na organização do ciclo ensino-aprendizagem-avaliação.

AVALIAÇÃO EXTERNA

Todos os anos são gastos alguns milhares de milhões de euros no sistema educativo português e, para importantes sectores da nossa sociedade, persiste uma insatisfação quase crónica face à qualidade do serviço que é prestado pelo sistema em geral e pelas escolas em particular. Sucessivas equipas governativas têm procurado contrariar este mal-estar da sociedade em relação ao sistema educativo. Infelizmente, apesar dos muitos esforços desenvolvidos, as equipas políticas acabam por se ver mais enredadas na gestão quase casuística dos múltiplos e corporativos interesses em presença, do que na concretização de medidas estruturantes subordinadas a uma agenda própria, decorrente de um plano estratégico para democratizar, modernizar e melhorar substancialmente a qualidade do sistema educativo. Entre os esforços que se têm realizado para reformar e reestruturar o sistema, contam-se os que se relacionam com o desenvolvimento do papel da avaliação na melhoria da qualidade do ensino e das escolas. Seguramente, nos últimos anos, largas dezenas de milhões de euros têm sido destinadas à avaliação externa, nomeadamente aos exames nacionais do ensino secundário, às provas de aferição, à participação nos estudos internacionais (TIMSS e PISA) e, a um outro nível, à avaliação integrada das escolas da responsabilidade da Inspeção Geral da Educação.

Em vários países do mundo, nomeadamente nos Estados Unidos da América, em Inglaterra, na Holanda, em França, na Austrália e nos países escandinavos, após muitos anos de exames nacionais e de resultados consistentemente modestos, têm surgido críticas bastante fortes às medidas utilizadas para monitorizar o desempenho dos alunos e para avaliar os respectivos sistemas. Em geral, tais críticas referem-se quer à forma quer ao conteúdo das medidas tradicionais, vulgo exames ou testes nacionais, que avaliam uma amostra muito reduzida dos domínios do currículo e, por isso, não avaliam muitos resultados significativos das aprendizagens dos alunos. Por outro lado, dizem os críticos, acabam por ter efeitos nefastos sobre o currículo, empobrecendo-o, sobre o ensino, demasiado condicionado pelo que «sai no exame», sobre o desenvolvimento de competências de resolução de problemas por parte dos alunos e sobre as decisões políticas. A um outro nível, têm sido muito debatidas na literatura as características dos exames nacionais, o seu papel na avaliação da qualidade dos sistemas educativos, as suas vantagens, as suas relações com a prestação de contas, as suas principais funções, a sua validade e fiabilidade, o problema da correcção, particularmente o da fiabilidade entre correctores, assim como questões de equidade, de comparabilidade e de ética. Tudo questões ainda muito pouco analisadas e debatidas entre nós.

As provas de aferição também se integram na avaliação externa e têm sido aplicadas aos alunos dos anos terminais dos ciclos da educação básica desde o ano lectivo de 1999/2000, nas disciplinas de Matemática e de Português. Importa analisar o seu papel na avaliação das aprendizagens e na avaliação do trabalho desenvolvido nas escolas assim como discutir as perspectivas que se poderão desenhar tendo em conta o anunciado início dos exames externos para todos os alunos do 9.º ano de escolaridade.

Finalmente, tal como acima referido, faz-se uma discussão acerca dos estudos internacionais de avaliação das aprendizagens.

Exames Públicos Nacionais

De há cerca de 150 anos para cá todos os países têm experimentado um desenvolvimento significativo, nalguns casos verdadeiramente notável, dos sistemas educativos. Ao longo destes anos, muitos países conseguiram que todos os seus cidadãos em idade escolar viessem, efectivamente, a frequentar a escola. Outros têm progredido significativamente em relação a esse objectivo, comum a todas as sociedades democráticas. Como consequência, na maioria dos países, assistimos a uma significativa expansão de todos os níveis de ensino. Começando, naturalmente, pela educação básica, os sistemas quiseram, e querem, que todos os seus alunos frequentem formações de nível secundário. A este crescimento do ensino secundário correspondeu um aumento muito significativo das ofertas ao nível do ensino superior que, em muitos casos, não foi suficiente para responder ao brutal aumento da procura.

Os países lidaram com estas situações de formas diferentes mas, em geral, utilizaram exames, ou outro tipo de provas, como forma de «regular» situações como a que sucedeu em Portugal, onde era muito maior o número de alunos a querer ingressar no ensino superior do que o número de vagas que as instituições podiam disponibilizar. Os exames assumiram, assim, uma função marcadamente selectiva que, eventualmente, poderá atenuar-se, ou mesmo desaparecer, se o sistema, entretanto, se tornar mais equilibrado e mais

aberto. Nesse caso, a função de certificação das aprendizagens dos alunos ganharia maior relevância do que a função da sua selecção.

É evidente que, como veremos, há outras questões, para além das suas funções, que devem ser analisadas relativas aos exames. Algumas dessas questões são abordadas nas páginas que se seguem.

Breve bosquejo histórico dos exames

Segundo vários autores (e.g., Keeves, 1994; Kellaghan e Madaus, 2003) os exames públicos nacionais parecem ter-se iniciado há cerca de 2500 anos na China, tendo em vista a selecção de pessoal militar e de funcionários públicos e, simultaneamente, evitar que a selecção fosse feita com base em laços familiares, ou outro tipo de afinidades, dos candidatos com os diferentes dignitários dos poderes instituídos.

Foram os jesuítas que, no século XVI, trouxeram os exames para a Europa, introduzindo-os nas suas escolas. Mas só no século XVIII, em 1748, a administração de exames em larga escala começou a ser utilizada na Europa, mais propriamente na antiga Prússia (Alemanha). O objectivo continuava a ser o de seleccionar funcionários públicos. Em França, após a revolução, em 1793, os exames são também adoptados. Só no século XIX o Reino Unido decide adoptar os exames como meio de seleccionar funcionários públicos para um império então em plena expansão. Nos Estados Unidos, os exames são introduzidos em 1883, ainda com o objectivo de seleccionar funcionários para o governo, depois de começarem a ser utilizados na cidade de Boston em 1845 (Kellaghan e Madaus, 2003).

No século XIX, ainda na Europa, as universidades acabam por ter um papel relevante na expansão dos exames públicos administrados em larga escala, pois começaram a utilizar muitos dos sistemas de exames existentes para seleccionar os seus alunos. É o caso do *Abitur*, na Alemanha, que era um exame de certificação usado na escola média alemã desde 1788 e que veio a ser utilizado como exame de qualificação para entrada nas universidades. É também o caso do *Baccalaureat*, introduzido em França por Napoleão, em 1808, que também era utilizado para acesso a certos cargos públicos e ao mercado de trabalho em geral. No Reino Unido, a Universidade de Londres inicia os seus exames – *matriculation examinations* – em 1838 (Kellaghan e Madaus, 2003).

Ao contrário de muitos países na Europa, os Estados Unidos não tinham, nem hoje têm, um sistema de exames públicos nacionais. Há exames públicos externos obrigatórios em alguns estados, nomeadamente o famoso *Regents examinations* no Estado de Nova Iorque, instituídos em 1860 e que ainda persistem. A partir das primeiras décadas do século XX, os norte-americanos, baseados nos princípios da psicometria, iniciaram o desenvolvimento de uma abordagem substancialmente diferente da dos europeus no que se referia à avaliação dos alunos através de provas em larga escala. Caracterizavam este tipo de avaliação a utilização predominante de testes objectivos de escolha múltipla e uma standardização de procedimentos de modo a que todos os candidatos, supostamente, realizassem os exames nas mesmas condições.

Presentemente, de acordo com Kellaghan e Madaus (2003) e Keeves (1994), praticamente todos os países do mundo têm um qualquer sistema de exames. Inclusivamente países que não tinham tradição de exames públicos externos, como é o caso de muitos países da ex-União Soviética e da Europa Oriental, estão neste momento a introduzi-los, particularmente no final do ensino secundário.

Em 1993, o Despacho n.º 338 do Ministério da Educação institui os exames nacionais no final do ensino secundário em Portugal, com funções de certificação e de selecção no acesso ao ensino superior. Assinale-se que o sistema educativo português estava, havia cerca de 20 anos, sem quaisquer exames ou qualquer tipo de avaliação externa com os efeitos definidos no despacho acima referido.

Características gerais dos exames

Analisando o que se passa em vários sistemas educativos, verifica-se que, nuns casos, há sistemas de exames ao nível das escolas, noutros, ao nível regional e, em muitos, ao nível nacional. Em certos sistemas educativos os exames têm apenas a função de certificação, noutros servem para certificar e para seleccionar, como é o caso dos nossos exames do ensino secundário, e alguns apenas se destinam a seleccionar, como é o caso de exames destinados a seleccionar alunos para certas vias ou percursos de nível secundário. Nuns sistemas educativos há alguma combinação, para efeitos de certificação ou de selecção, entre as classificações obtidas pelos alunos nos exames nacionais e as que são obtidas nas escolas. Noutros sistemas a certificação e/ou a selecção é baseada apenas nas classificações externas, obtidas nos exames, ou nas classificações internas, obtidas na chamada avaliação contínua das escolas.

Em todo o caso, é possível identificar um conjunto de características comuns a um elevado número de países. Por exemplo, Kellaghan e Madaus (2003), identificam as seguintes:

1. Os exames são externos, isto é, são preparados e controlados por uma ou mais entidades externas às escolas a que os alunos pertencem;
2. A administração dos exames é normalmente controlada pelo governo ou, no mínimo, por si supervisionada;
3. Os exames são construídos a partir dos conteúdos constantes nos currículos. Em geral, isto significa que a ênfase é mais no conhecimento dos conteúdos do que, por exemplo, na sua integração e mobilização para resolver situações problemáticas;
4. As provas de exame são iguais para todos os alunos, administradas segundo procedimentos standardizados, fora do ambiente normal da sala de aula e realizados por um grande número de alunos;
5. A maioria dos exames tem várias funções, tais como certificar, controlar, seleccionar ou motivar;
6. Normalmente, o conteúdo, os critérios de correcção e os resultados dos exames são tornados públicos.

Apesar destas características comuns a um elevado número de sistemas, há, como é natural, um conjunto de diferenças, das quais destaco as seguintes:

1. Em geral, os exames ocorrem no final do ensino secundário, mas há países que utilizam exames, pelas mais variadas razões, a níveis mais elementares da escolaridade, normalmente quando os alunos têm cerca de 16 anos ou, em muitos casos, quando estão no final da escolaridade obrigatória;
2. O número de entidades envolvidas no processo de exames varia de país para país. Por exemplo, no nosso caso, há pelo menos sete instituições ou entidades envolvidas:
 - O Gabinete de Avaliação Educacional (GAVE), que elabora as provas e os critérios de correcção;
 - O Departamento do Ensino Secundário (DES) que é responsável pela administração das provas, pela coordenação do processo de correcção, pela recolha, tratamento e difusão da informação e pelo processo de revisão de provas solicitadas pelos alunos;
 - O Júri Nacional de Exames do Ensino Secundário (JNE), que elabora o respectivo regulamento e controla mecanismos operacionais para o fazer observar;
 - As Direcções Regionais de Educação (DRE), que cooperam com as entidades centrais em questões de natureza logística e relacionadas com o processo de administração dos exames;
 - A Inspecção Geral da Educação (IGE), que zela, junto das escolas, para que os procedimentos estejam conforme os regulamentos, garantindo que os exames decorram em condições de igualdade para todos os alunos;
 - A Editorial do Ministério da Educação, que imprime as provas e assegura a sua distribuição por todas as escolas do país onde se realizam exames; e
 - As forças de segurança – PSP e GNR – que colaboram com a Editorial para garantir que a distribuição seja feita sem problemas;
3. Há países em que o peso dos exames para efeitos de certificação ou de selecção é de 100%, ou muito elevado (mais do que 70%). Por exemplo, em Inglaterra o peso dos exames no cômputo da classificação final dos alunos é de 80%, contra 20% da avaliação realizada nas escolas. Em Portugal, para efeitos de certificação, o peso dos exames é de 30% e o da avaliação interna é de 70%. Mas, para efeitos de selecção das entradas no ensino superior, o peso das disciplinas específicas de exame (normalmente duas) é de 50% contra outros 50% da avaliação interna. Refira-se que só num reduzido número de países europeus, como é o caso da Irlanda, da França e da Inglaterra, a avaliação no final do secundário é totalmente ou quase totalmente externa. Noutros países, há exames que são administrados e corrigidos pelos professores nas escolas com algum sistema de moderação externa;

4. O número de exames que os alunos têm que fazer também varia. Por exemplo, em Itália os alunos fazem dois exames, na Alemanha e em Inglaterra fazem três, em Portugal entre quatro e cinco e em França, na Irlanda e na Holanda fazem mais de seis exames;
5. O formato dos exames também pode variar sensivelmente quer quanto ao número de partes que o constituem quer quanto à natureza e número das questões. Em muitos sistemas educativos as questões tipo ensaio têm sido predominantes mas também fazem uso de questões de resposta curta e de questões objectivas (vulgo escolha múltipla). Há exames que incluem tarefas de desempenho para avaliação de competências práticas e orais.

Verifica-se, assim, que há uma variedade de modalidades, de funções, de formatos, de propósitos e de formas de controlo, de acordo com as políticas educativas dos diferentes países e das prioridades ou objectivos que definem.

Por outro lado, são conhecidos e estão amplamente documentados, os diferentes tipos de efeitos ou impactos que os exames podem ter aos mais variados níveis, tais como:

1. Nas vidas pessoais, sociais e académicas dos alunos;
2. Nas formas como as escolas e os professores se organizam e desenvolvem o currículo;
3. Naquilo que é ensinado e como é ensinado;
4. Naquilo que é avaliado e como é avaliado;
5. Na credibilidade social dos sistemas educativos.

Em função destes e doutros impactos dos exames, há, por sua vez, um conjunto de questões que, neste contexto, devem ser consideradas e discutidas como é o caso da validade, da fiabilidade, da comparação, da correcção ou da equidade. Consequentemente, surgem imediatamente problemas interessantes para analisar e discutir que se podem traduzir pelas seguintes questões:

1. O que é que os exames, de facto, avaliam?
2. Que currículo é avaliado pelos exames?
3. Qual a consistência dos resultados dos exames?
4. Os exames serão igualmente justos para todos os alunos?
5. Haverá grupos de alunos particularmente favorecidos ou particularmente desfavorecidos pelos exames?
6. A correcção tem níveis aceitáveis de consistência? Isto é, há fiabilidade entre os correctores? Como se poderá melhorá-la ou garanti-la?
7. A análise, apresentação e divulgação dos resultados é aceitável?
8. Há um plano estratégico, consistente e deliberado de investigação associado à realização dos exames?
9. Que consequências se retiram dos resultados dos exames e da análise dos dados?

Estas e outras questões integram as preocupações mais constantemente expressas na literatura e que aqui se procuram analisar e discutir, sem deixar de enunciar vantagens e desvantagens de um tipo de avaliação que, nos tempos que correm, parece estar em plena ascensão nos sistemas educativos.

Funções dos exames

Na maioria dos países, os exames desempenham um conjunto variado de funções, algumas das quais muito associadas à função sumativa da avaliação. As funções normalmente referidas são as seguintes:

1. *Certificação*, que cumpre o papel de comprovar que um dado candidato, ao fim de um dado período de tempo, demonstrou possuir um dado conjunto de aprendizagens;
2. *Seleção*, que está muito associada à função de certificação e que controla a progressão escolar dos alunos, nomeadamente o acesso ao ensino superior;
3. *Controlo*, que permite que os governos, através de exames standardizados por si controlados, assegurem que conteúdos semelhantes sejam leccionados nas escolas do país de acordo com o previsto no currículo nacional;

4. *Monitorização*, muito associada à prestação de contas e que, em alguns países, consiste na utilização dos resultados dos exames para «pedir contas» às escolas e aos professores. Normalmente promove-se a publicação de *rankings* das escolas de acordo com o seu desempenho nos exames nacionais;
5. *Motivação*, é uma das funções principais quando os exames não têm qualquer efeito na vida escolar dos alunos, mas, no caso em que têm efeitos importantes no seu progresso escolar, podem ser desmotivadores, particularmente junto dos alunos que consideram os exames difíceis.

A função de monitorização, como se viu, aparece na literatura associada à prestação de contas e à publicação dos *rankings* das escolas. A este propósito julga-se importante discutir aqui algumas razões que desaconselham a prática dos *rankings*.

De modo geral, as comparações que se possam fazer entre escolas tendem a não ter em devida conta os seus pontos de partida; isto é, não consideram o tipo de alunos que a frequentam, nem as qualificações dos professores e dos técnicos e funcionários da escola, nem os recursos materiais ou as condições físicas da escola. Além disso, as características da comunidade onde a escola está inserida também têm de ser devidamente consideradas. É evidente que os professores e as escolas têm sempre responsabilidades no desenvolvimento das aprendizagens dos alunos, mas poderemos estar a cometer uma injustiça se julgarmos uma escola baseados exclusivamente numa classificação que, por vezes, pode resultar de uma agregação de dados sem muito sentido.

Quando, num dado país, as escolas são confrontadas anualmente com um *ranking* oficial com base no desempenho dos seus alunos em exames nacionais, é natural que comecem a delinear estratégias defensivas que lhes possam garantir um bom lugar no dito *ranking*. Uma delas é a de «convidar» os alunos a desistirem, anulando as respectivas matrículas ou, pura e simplesmente, não levar os alunos a exame. Dito de outra maneira, certas escolas poderão ter critérios, mais ou menos explícitos, bastante «apertados» através dos quais muito poucos alunos conseguem aceder aos exames. Pode sempre perguntar-se que significados atribuir ao facto de uma escola que, no início do ensino secundário, tem 200 alunos dos quais apenas 50 acabam por se apresentar a exame e está num lugar cimeiro do *ranking*, e ao facto de uma escola com um lugar modesto no *ranking* que tinha inicialmente 200 alunos iniciais dos quais 185 foram a exame. Qual é a escola boa? A que reprovou 150 alunos? A escola má é a que apenas reprovou 15 alunos porque, eventualmente, seria muito permissiva? Há naturalmente muitas variáveis que poderiam ser consideradas. O que importa sublinhar é que não se trata de matéria pacífica e que, mais uma vez, se poderão estar a cometer injustiças acerca dos juízos que se façam sobre certas escolas.

Tem-se constatado, por outro lado, que há tantos *rankings* quantos os que quisermos. De facto, tudo depende dos resultados da escola que se utilizarem para produzir o *ranking*. Não é fácil encontrar uma variável que possa traduzir a multidimensionalidade dos problemas a considerar e, sobretudo, uma variável que faça sentido e que tenha real significado. Entre nós têm surgido ordenações diferentes das escolas conforme os critérios utilizados pelos seus autores.

Apesar de facilmente se perceber que há erros vários que se cometem na elaboração de *rankings* ou, no mínimo, pressupostos incorrectos ou procedimentos muito pouco credíveis, a verdade é que, em geral, não são mencionados nem tidos em conta pelas autoridades, pela sociedade ou pela comunicação social.

A elaboração e publicação de *rankings* pode ter outros efeitos indesejáveis como o de estigmatizar escolas que eventualmente até poderão estar a fazer um melhor trabalho do que escolas melhor classificadas no *ranking*, desmoralizando os professores e restantes colaboradores da escola e a própria comunidade em que a escola se integra.

Apesar de, como se disse, os resultados dos alunos estarem necessariamente relacionados com o que lhes é ensinado e como lhes é ensinado, é redutor e precipitado estar a avaliar uma escola e os seus professores com base exclusiva em tais resultados. A investigação já há muitos anos que tem mostrado isso mesmo, assim como a dificuldade e a complexidade de se conceberem sistemas de avaliação dos professores e das escolas que sejam exequíveis e que contribuam efectivamente para a melhoria dos sistemas educativos. Uma coisa parece certa, avaliar escolas e professores apenas com base nos resultados dos alunos é uma simpática tentação mas não deverá ser mais do que isso (Johnson, 2003; Pearlman e Tannembaum, 2003; Sanders e Davidson, 2003; Stufflebeam, 2003).

A função de monitorizar as escolas através dos exames tem naturalmente o seu papel e não tem necessariamente que passar pela publicação de *rankings* de escolas. É necessário considerar todas as faces do

problema e, sobretudo, quando se trata de atribuir responsabilidades, ter o cuidado de considerar todas as instituições, todas as pessoas que intervêm no processo educativo. Seria estranho que alguém pudesse, ou quisesse, ficar de fora.

Considerações sobre o tipo de questões

Como já atrás se fez referência, os exames podem incorporar uma variedade de questões tais como:

- Tarefas ou problemas;
- Ensaios;
- Resposta curta;
- Escolha múltipla.

As questões mais abertas, tipo problemas, tarefas ou ensaios, são fortemente recomendadas pelos autores que se enquadram no paradigma da avaliação alternativa pois, entre outras, apresentam as seguintes vantagens:

1. Permitem analisar os processos e estratégias utilizados pelos alunos na resolução de problemas novos;
2. Permitem avaliar as capacidades dos alunos para integrarem, relacionarem, aplicarem e organizarem conhecimentos;
3. Permitem avaliar as capacidades dos alunos para analisarem, sintetizarem e avaliarem a informação constante numa grande diversidade de textos ou de situações problemáticas;
4. Permitem avaliar aspectos originais e criativos do pensamento dos alunos.

As questões deste tipo, no contexto de um exame nacional, com efeitos no progresso escolar dos alunos e/ou na certificação, apresentam os seguintes problemas:

1. São normalmente difíceis de elaborar;
2. Demoram bastante tempo a ser corrigidas;
3. Os critérios de correcção são difíceis de elaborar;
4. A fiabilidade entre os correctores tende a baixar;
5. A fiabilidade do exame tende a baixar.

Apesar de todos estes problemas deverem ser igualmente considerados numa avaliação em larga escala, com os efeitos dos exames em vigor em Portugal, talvez a mais preocupante seja a fiabilidade entre os correctores. Sobretudo porque pode questionar de forma mais ou menos séria a equidade dos exames. Sabe-se que há procedimentos que podem ajudar a diminuir este problema, alguns dos quais têm sido postos em prática pelas entidades responsáveis, como é o caso do maior cuidado que tem sido posto na elaboração de critérios de correcção, de algumas acções de formação para correctores ou a promoção de reuniões entre correctores para aferição de critérios. Mas também se sabe que há aqui um longo caminho a percorrer. Perante esta situação, parece importante que se criem mecanismos mais apertados de controlo da correcção e que, através de amostras aleatórias de provas, se determine quais são os níveis de fiabilidade das correcções. É uma questão relevante que não pode deixar de ser tratada com a maior transparência.

É evidente que a solução não é eliminar as perguntas abertas relativas à resolução de problemas ou a outro tema qualquer. Por razões óbvias é desejável e indispensável que os exames mantenham esse tipo de questões.

As perguntas de resposta curta e de escolha múltipla também são utilizadas entre nós nos exames nacionais do ensino secundário, pelo menos em parte das provas de algumas disciplinas, nas provas de aferição do ensino básico desde o ano 2000 e também nas provas dos estudos internacionais em que o nosso país tem participado.

Os testes ou perguntas de escolha múltipla, uns e outras largamente utilizados em exames nacionais em muitos países do mundo surgem nos Estados Unidos da América, em 1914, para responder a duas dificuldades: a) o número crescente de alunos no sistema educativo associado à chegada de milhões de imigrantes aquele país; e b) o facto de se detectarem grandes discrepâncias na correcção de exames em que as perguntas eram do tipo ensaio. Na verdade, investigações conduzidas nos finais do século XIX e princípios do século XX mostraram que as classificações atribuídas aos testes com perguntas abertas, ou tipo ensaio, variavam muito

e dependiam de quem os corrigia. Os sistemas de exames ou de testes foram então considerados pouco eficientes, ou mesmo ineficientes, e excessivamente subjectivos. Os testes de escolha múltipla acabaram por se afirmar no contexto norte-americano, particularmente após a descoberta dos instrumentos de leitura óptica das respostas nos anos 50, mas rapidamente se foram disseminando por muitos sistemas educativos em todo o mundo. As razões para o seu aparente sucesso estavam relacionadas com o facto de permitirem examinar um grande número de alunos em pouco tempo e, sobretudo, porque a fiabilidade da correcção era total dado que as perguntas, supostamente, eram todas objectivas.

Kellaghan e Madaus (2000) referem que os testes de escolha múltipla são utilizados para avaliar aprendizagens adquiridas pelos alunos pelas seguintes razões:

1. A correcção é objectiva; num dado teste as classificações serão sempre as mesmas independentemente de quem o corrija;
2. Os testes são muito eficientes porque podem ser facilmente administrados a muitos alunos e serem corrigidos rapidamente, sem erros e de forma muito barata devido à utilização de instrumentos de leitura óptica. A administração é muito simples e não exige formação especial dos intervenientes no processo;
3. Os testes de escolha múltipla podem abranger uma parte substancial do domínio a avaliar em relativamente pouco tempo. Por exemplo, em 40 minutos os alunos podem responder a cerca de 50 questões;
4. Os testes de escolha múltipla são normalmente fiáveis; isto é, os resultados tendem a ser consistentes se os testes, ou alguma versão considerada equivalente, forem administrados, em tempos diferentes, a populações comparáveis ;
5. Os testes de escolha múltipla, talvez porque a sua correcção é objectiva e são largamente utilizados, são considerados como boas e credíveis fontes de informação;
6. As classificações utilizadas nos testes de escolha múltipla são facilmente compreendidas pelas pessoas. (Kellaghan e Madaus, 2000, p.120)

Os mesmos autores listam também as seguintes limitações associadas aos testes de escolha múltipla:

1. Os testes de escolha múltipla são muitas vezes ambíguos. Uma redacção de um item que pareça clara para quem constrói o teste pode confundir quem tem que o resolver;
2. Os testes de escolha múltipla proporcionam muito pouca informação diagnóstica acerca dos alunos. Não fornecem informação acerca das razões que estão na origem de eles acertarem, ou não, nas questões e acerca dos processos utilizados para lhes responderem;
3. Os testes standardizados de escolha múltipla fornecem muita pouca informação que possa ser directamente utilizada pelos professores para orientar ou melhorar o ensino;
4. Os testes de escolha múltipla, em geral, não nos fornecem informação acerca dos processos mais complexos de pensamento dos alunos. É mais fácil construir itens de escolha múltipla para avaliar o conhecimento de factos, do que itens que tenham a ver com processos mentais mais complexos que se desenvolvam em múltiplas etapas;
5. Se os testes de escolha múltipla tiverem efeitos sobre a progressão académica dos alunos, seja para progredir para o ano seguinte ou para concluir o ensino secundário, podem exercer uma influência negativa no currículo. (Kellaghan e Madaus, 2000, p.120)

São múltiplas as consequências que resultam da utilização de um teste com determinadas características. O princípio de procurar diversificar o tipo de perguntas ou de itens, como me parece que tem sido feito entre nós, tem virtualidades importantes. A ideia deve ser a de procurar equilibrar as exigências de validade, de fiabilidade, de equidade e de capacidade de comparação que têm que existir neste tipo de provas com as exigências de um ensino e de uma aprendizagem em que a resolução de problemas, a relação e integração de saberes ou as competências metacognitivas das pessoas têm um papel determinante.

Sabemos através da experiência inglesa com a *Assessment of Performance Unit* – APU, (Torrance, 2003), no início dos anos 90, que, em princípio, quanto maior for a complexidade das tarefas (e.g., tarefas abertas, resolução de problemas), mais difícil é incluí-las em avaliações em larga escala. Ou seja, tendencialmente, quanto maior for a escala de abrangência dos exames e dos seus efeitos sobre o progresso e a certificação dos alunos, mais «simples» (mais objectivo) deverá ser o tipo de itens. E isto porque assim os alunos tenderão a ser tratados todos da

mesma maneira na correcção das provas. Há uma tensão entre os propósitos formativos e avaliativos do sistema nacional de exames e a relação entre a escala (larga, média, pequena) e a complexidade e objectividade das perguntas (objectivas, resposta curta, ensaio ou questões abertas).

Muitas coisas importantes estão em jogo: «estreitamento» ou não do currículo, equidade do sistema, indução de boas ou más práticas nas escolas ou, talvez surpreendentemente para alguns, a real possibilidade de se baixarem os níveis de aprendizagem dos alunos. São coisas demasiado sérias para não se discutirem aberta e responsabilmente entre nós.

Equidade

A questão da equidade no sistema educativo e, em particular, na avaliação e nos exames nacionais não tem sido uma questão debatida entre nós, pelo menos como tem sido discutida, investigada e reflectida noutros países europeus (e.g., Apple, 1989; Gipps e Murphy, 1994). Normalmente a discussão tem-se circunscrito à necessidade de criar «igualdade de oportunidades» para todos os alunos e, particularmente, para os que pertencem a grupos que carecem de necessidades educativas especiais e a outros normalmente designados por «minoritários». E isto pode ter a ver com a origem sócio-económica, a origem étnica, a filiação religiosa, o género, a língua materna ou com qualquer outra característica. Em geral, procura-se assegurar igualdade de oportunidades, com sistemas, mais ou menos eficientes, de apoios e complementos educativos. A perspectiva que, em muitos países, tem sido utilizada é a de proporcionar igualdade de acesso aos recursos educativos e aos diferentes percursos de educação e de formação previstos nos sistemas educativos. Trata-se, sem dúvida, de uma abordagem importante ao problema, mas que tem sido considerada insuficiente, porque os diferentes grupos de alunos, por exemplo, os rapazes e as raparigas ou alunos de etnias diferentes têm experiências fora da escola que podem ser significativamente diferentes. Ou seja, o modelo procura compensar as desvantagens de certos grupos de alunos garantindo-lhes acesso ao tipo de «serviço» a que todos os seus colegas têm acesso. Seria interessante desde que algumas coisas pudessem mudar na forma como se organiza o ensino ou se envolvem estes alunos nas tarefas e na comunidade escolar. O que acontece é que normalmente proporciona-se igualdade de acesso aos meios mas tudo o resto se mantém inalterado. Por isso, são poucos os que eventualmente superam as suas dificuldades.

Outra abordagem ao problema é a que se centra na igualdade de resultados que deve traduzir, em princípio, o acesso à igualdade de oportunidades e que, necessariamente, recentra a discussão nos resultados dos diferentes grupos de alunos.

Outras concepções consideram a equidade como o justo e adequado acesso à educação por parte de todos os alunos e, conseqüentemente, como a real possibilidade de todos os grupos obterem resultados «razoavelmente semelhantes» em provas de avaliação, como é o caso dos exames nacionais obrigatórios. Desta forma, o princípio da equidade em educação significaria que todos os alunos teriam acesso a uma educação e a uma formação semelhantes ou iguais e que, por consequência, teriam resultados semelhantes nas avaliações. Nestas condições, a igualdade de resultados acaba por constituir a base mais sólida para a igualdade de oportunidades para todos os grupos de alunos que saem das escolas.

Gipps e Murphy (1994), por seu lado, sustentam que a igualdade de resultados não é propriamente uma finalidade apropriada, porque grupos diferentes podem ter experiências, interesses, motivações, características, qualidades e pontos fortes e fracos muito distintos. Para estas autoras, a equidade na avaliação implica que as suas práticas e a interpretação dos resultados seja justa, adequada e equilibrada para todos os grupos.

Ainda noutras concepções sustenta-se a ideia de que a igualdade de oportunidades não está associada à opressão dos grupos e às suas desvantagens, mas antes à preocupação em garantir que todos os alunos possam escolher individualmente, numa perspectiva de «mercado livre», o que querem fazer das suas vidas escolares. Facilmente se percebe que esta perspectiva centra nos alunos a responsabilidade pelas suas desvantagens ou dificuldades de aprendizagem ao mesmo tempo que isenta as escolas e a comunidade de assumirem quaisquer papéis porque «as oportunidades estão lá». O aluno só tem que escolher o que vai mais ao encontro das suas necessidades, motivações e capacidades!

Como reacção a estas perspectivas surgem outros que fazem passar o debate da equidade por questões curriculares tão relevantes como as seguintes:

1. Que tipos de saberes, atitudes e capacidades são ensinados a quem?
2. Por que é que os saberes, atitudes e capacidades são ensinados de certas formas a determinados grupos particulares?

3. De que formas conseguimos que as histórias e culturas dos diferentes grupos de alunos sejam responsabilmente integradas nos processos de ensino e aprendizagem?

É tendo em conta o que acima se discutiu que, num sistema nacional de exames com efeitos na vida académica, pessoal ou social dos alunos, como é o caso dos exames nacionais do ensino secundário em Portugal, as questões da equidade devem ser consideradas. Dir-se-á que os exames nacionais asseguram a equidade quando os jovens que, à partida, desenvolveram o mesmo tipo de aprendizagens, obtêm resultados iguais ou, pelo menos, muito semelhantes. Se o sistema não assegura a equidade, então poderá haver grupos de estudantes que são beneficiados, ou prejudicados, em relação a outros. Fica assim a porta aberta para que certos grupos de alunos fiquem à frente ou atrás de outros por razões que nada têm a ver com o tipo de conhecimentos ou aprendizagens testados no exame.

As questões de equidade têm que ser equacionadas no sistema educativo em geral, pois é no seu seio que as diferenças de tratamento, as discriminações de ordem vária e a falta de oportunidades para aprender podem ser sentidas por certos grupos de estudantes e não por outros. Por estas razões, há investigadores que consideram que o problema da equidade nos sistemas educativos é um problema central e que, sob muitos pontos de vista, constituem uma justificação para os exames externos (Kellaghan e Madaus, 2003).

O que parece poder acontecer é que, em exames nacionais, haja algumas parcialidades que acabem por afectar, prejudicando ou beneficiando, certos grupos de alunos – rapazes, raparigas, grupos étnicos, grupos linguísticos, candidatos provenientes de meios urbanos ou rurais ou candidatos de escolas públicas ou privadas. A investigação tem mostrado que as diferenças nas classificações dos alunos estão normalmente associadas ao estatuto sócio-económico dos candidatos, à sua origem étnica ou ao seu género, continuando por clarificar se tais diferenças são devidas ao exame ou ao que os alunos efectivamente sabem e são capazes de fazer.

De acordo com Bolger e Kellaghan, Mathews e Stage (citados em Kellaghan e Madaus, 2003), há um resultado consistente num número de países que atribui ao método utilizado para avaliar as aprendizagens as diferenças entre géneros. Assim, os rapazes obtêm consistentemente melhores resultados do que as raparigas em testes standardizados de múltipla escolha enquanto que as raparigas obtêm melhores resultados em testes que contêm perguntas abertas ou do tipo ensaio.

Há um conjunto de factores que deve ser considerado na abordagem sistemática que é necessário fazer-se à questão da equidade. Já vimos que o estatuto sócio-económico dos alunos, o seu género, o grupo étnico e/ou linguístico a que pertencem, a situação da localidade de proveniência (urbana ou rural) ou tipo de escola (pública ou privada) são factores a ter em conta.

Outra questão que tem que ser devidamente enquadrada é a que se refere aos candidatos com necessidades educativas especiais. À medida que o número de alunos que fazem exames aumenta, surgem cada vez mais alunos que necessitam de algum tipo de assistência, que vai desde provas com dimensões e arranjos especiais, para alunos amblíopes, até proporcionar um professor ou um técnico de apoio para alunos com problemas motores, ou meios tecnológicos especiais para crianças ou jovens paraplégicas ou tetraplégicas. O sistema deve assegurar que nenhuma criança com necessidades educativas especiais deixe de ter as condições de realização do exame que lhe permitam mostrar o que sabe e o que é capaz de fazer.

Validade e fiabilidade dos exames

A validade e a fiabilidade são duas das principais características psicométricas da avaliação.

Tradicionalmente diz-se que um teste é válido se avalia realmente aquilo para que foi construído. Se um teste não avalia o que é suposto avaliar então a sua utilização é enganadora.

Em geral, a literatura sobre esta matéria refere os seguintes tipos principais de validade:

- a validade de previsão (em que medida um teste é um bom indicador de desempenhos futuros da pessoa que o resolve);
- a validade de conteúdo (em que medida é que um teste contém uma amostra significativa do conteúdo relevante do domínio ou domínios que foram ensinados);
- a validade concorrente (em que medida os resultados do teste se correlacionam com os resultados doutro teste ou de outra avaliação das mesmas aprendizagens);
- a validade de critério (em que medida o teste permite prever o desempenho relativo a um dado critério);
- a validade de constructo (até que ponto o teste é uma medida adequada do constructo; isto é, da competência subjacente que está a ser avaliada).

Apesar de todos estes tipos de validade, a verdade é que os investigadores normalmente se referem a um ou dois tipos, com particular relevância para a validade de conteúdo.

De acordo com Gipps (1994), a validade é hoje encarada como um conceito unitário em que o constructo (a competência ou a aprendizagem subjacente) é o tema unificador. Trata-se de uma interessante e apaixonante discussão que, no fundo, considera que a validade de um teste ou de um exame não tem a ver só com o teste em si mesmo, ou com os resultados que produz, mas também com a utilização que se faz desses resultados e com as consequências das decisões avaliativas. É matéria que se abordará detalhadamente na versão definitiva deste texto.

A fiabilidade de um teste ou de um exame nacional tem a ver com a consistência dos seus resultados. Ou seja, para analisarmos se um exame é fiável temos que quantificar em que medida o desempenho dos examinandos se mantém sensivelmente o mesmo, se resolverem o exame em tempos ou ocasiões diferentes. A fiabilidade tem uma relevância acrescida quando os exames têm a sua função de selecção muito marcada, como é o caso do que se passa entre nós, pois, como vimos, os exames servem para seleccionar alunos para o ensino superior. É por isso que se afirma que a principal questão é a da replicação. Ou seja, se, na sequência de um dado exame e de uma resultante selecção de candidatos, tivesse lugar uma replicação desse mesmo exame, põe-se a questão de saber se os candidatos seleccionados se manteriam os mesmos ou não. Se sim, então o exame é considerado fiável e, em última análise, justo e equilibrado, tratando, com equidade, todos os candidatos. Se não, então temos um problema sério, pois o exame permite que a selecção de candidatos ao ensino superior possa estar dependente de outros factores que não os conhecimentos ou as reais aprendizagens adquiridas pelos candidatos.

Há inúmeros factores que podem influenciar a fiabilidade de um exame, alguns inerentes a qualquer tipo de avaliação e outros mais relacionados com avaliações em que há uma grande incidência de questões abertas. Analisemos então alguns desses factores.

Kellaghan e Madaus (2003) referem quatro factores que podem afectar a fiabilidade de um exame:

1. Os alunos podem ter desempenhos diferentes em momentos de resolução diferentes;
2. Os desempenhos dos alunos podem ser influenciados por condições externas ao próprio exame;
3. Os desempenhos dos alunos podem variar com a variação das questões que têm que resolver;
4. As correcções dos exames podem variar sensivelmente de corrector para corrector, principalmente em questões não objectivas, de resposta aberta.

Para diminuir, ou mesmo anular, estas «ameaças» à fiabilidade dos exames, o que normalmente se faz é standardizar as condições de administração; os critérios de correcção, que devem tão detalhados quanto possível; e os procedimentos dos correctores, para assegurar que os critérios e os padrões de correcção são uniformes. É também usual apresentar provas corrigidas, aos correctores, em que constam as explicações para as pontuações que se atribuem.

Como é evidente, quanto mais «apertadas» forem estas condições mais limitações acabam por surgir quanto ao tipo de tarefas e, portanto, de conhecimentos, que se acabam por poder avaliar.

Gipps e Stobart (2003) chamam a atenção para o facto das avaliações alternativas serem particularmente vulneráveis aos problemas de fiabilidade. Nomeadamente referem que é difícil administrar questões abertas de «alta fiabilidade» devido ao tempo que demoram a ser resolvidas. Assim, há limitações sérias quanto às inferências que se possam fazer acerca dos desempenhos dos alunos, porque nunca se poderão administrar muitas questões desta natureza num exame. Repare-se que Shavelson, Baxter e Gao, citados em Gipps e Stobart (2003), estimaram que, para se atingirem níveis aceitáveis de fiabilidade, seriam necessárias entre 10 e 23 questões daquele tipo! Relativamente à fiabilidade das correcções, Gipps e Stobart parecem mais optimistas, pois consideram que pode ser significativamente melhorada com processos de formação, critérios de correcção muito detalhados e tarefas standardizadas.

Gipps (1994), enumera as quatro técnicas clássicas para avaliar a fiabilidade de um teste:

1. Administrar o mesmo teste com alguns dias de intervalo e comparar os desempenhos obtidos em cada administração;

2. Administrar versões comparáveis do mesmo teste a amostras semelhantes de uma dada população e comparar os respectivos resultados;
3. Se só é possível proceder a uma administração ou se só há um teste disponível, divide-se aleatoriamente o teste em duas partes que se administram separadamente, comparando-se os resultados obtidos pelos alunos nas duas partes;
4. Determinar estatisticamente um coeficiente de consistência interna a partir de todas as correlações que se calculam a partir de todas as possíveis «divisões» do teste (por exemplo, calcular a correlação entre todos os resultados do item x e todos os resultados finais do teste).

Para melhorar a consistência das correcções, Gipps enumera um conjunto de processos de moderação externa:

1. Moderação estatística através da utilização de testes de referência ou técnicas de *scaling*;
2. Moderação por inspecção;
3. Moderação por um painel de revisão;
4. Moderação por consenso;
5. Moderação por grupos;
6. Moderação intrínseca;
7. Moderação através da acreditação de instituições.

Nem todos estes métodos são passíveis de utilizar num contexto de exames nacionais, pois alguns destinam-se mais a «moderar» avaliações internas das escolas.

Vantagens e desvantagens dos exames

Os exames podem ter certamente um conjunto de vantagens e, talvez, em parte, por essa razão, são aplicados em quase todos os países. Entre algumas vantagens que normalmente se associam aos exames, partindo do princípio que a sua concepção e elaboração tem qualidade pedagógica, educativa e formativa, poderemos indicar as seguintes:

1. Podem exercer um efeito moderador importante nas avaliações internas;
2. Podem induzir práticas inovadoras de ensino e de avaliação;
3. Podem contribuir para avaliar o sistema educativo e ajudar a melhorar a tomada de decisões a todos os níveis;
4. Podem ajudar as escolas a melhorar os seus projectos;
5. Podem dar indicações às escolas, aos professores e aos alunos acerca do que é importante ensinar e aprender.

Por outro lado, os exames também possuem desvantagens que não se podem ignorar:

1. Centram-se sobretudo nos conhecimentos académicos, prestando relativamente pouca atenção a conhecimentos mais úteis, relacionados com a vida real;
2. Condicionam os objectivos, as estratégias, o envolvimento e as atitudes dos alunos em relação às aprendizagens, nem sempre pelas melhores razões. Exames que usam perguntas objectivas e de resposta curta tendem a deixar de avaliar aprendizagens consideradas fundamentais, como é o caso da resolução de problemas;
3. Podem induzir práticas fraudulentas. Trata-se de uma questão séria em muitos países, onde a corrupção, a compra de cópias de enunciados de exames, a substituição de alunos que eram supostos fazer o exame ou a cumplicidade entre professores vigilantes e examinandos são práticas comuns;
4. Podem induzir as escolas a concentrarem os seus esforços nos alunos que têm mais possibilidades de ter sucesso nos exames,
5. Podem discriminar, em vez de integrar, alunos. Particularmente certos grupos pertencentes a minorias de qualquer natureza.

Avaliação Aferida

Em 1991 publicou-se o Despacho n.º 162/ME/91, que aprovava o sistema de avaliação dos alunos dos ensinos básico e secundário e que acabou por ser um primeiro passo para que os princípios e orientações constantes dos diplomas legais a partir de então passassem a ser mais consistentes com as exigências curriculares, com as formas de aprendizagem e com as necessidades de desenvolvimento democrático do sistema educativo. Se quisermos, foi um despacho que, até certo ponto, «alinhou» parte dos seus conteúdos com o «movimento» que, um pouco por todos os países mais desenvolvidos, se tem vindo a designar por *avaliação alternativa* (Gipps e Stobart, 2003), *avaliação autêntica* (Newmann e Archbald, 1992; Tellez, 1996) ou *avaliação educativa* (Gipps, 1994), de inspiração cognitivista e construtivista, em contraste com a avaliação dominante, inspirada nos modelos psicométricos de medição da inteligência e das aptidões.

O Despacho teve um processo de desenvolvimento que, pelo menos na altura, se poderia considerar inédito. De facto, a sua publicação definitiva foi precedida de uma discussão pública e de um estudo do tipo das sondagens que recolheu as opiniões (geralmente de tendência favorável) de cerca de 60000 professores de todos os ciclos de ensino acerca das «medidas constantes no projecto do sistema de avaliação dos alunos» (Fernandes, Ramalho e Lemos, 1991).

Uma das novidades foi a criação da avaliação aferida que, «para efeito de progresso escolar dá-se apenas no final do ensino básico e no final do ensino secundário». Ou seja, no final da educação básica e no final da educação secundária, a avaliação aferida teria efeitos ao nível da certificação dos alunos e ao nível do prosseguimento de estudos. Para efeitos do cálculo da classificação no final da educação básica, a Avaliação Aferida teria peso 1 e a «Avaliação Formativa» teria peso 2. A média da chamada Avaliação Formativa (Avaliação interna, da responsabilidade exclusiva dos professores) era ponderada através do número de horas semanais de cada disciplina. Em suma, para efeitos de classificação final e atribuição do respectivo diploma, a avaliação externa tinha um peso de cerca de 33% e a avaliação interna um peso de cerca de 66%.

As funções atribuídas pelo Despacho mencionado à avaliação aferida eram, no essencial, as seguintes:

1. Verificar o grau de cumprimento dos objectivos curriculares;
2. Controlar e avaliar a qualidade do sistema educativo;
3. Credibilizar socialmente os diplomas escolares;
4. Moderar externamente as classificações resultantes da avaliação interna.

O que se poderá nesta altura referir é que então se procurou compatibilizar um sistema de avaliação interna, que continha um conjunto de princípios e de orientações mais inspirados nas avaliações alternativas de inspiração construtivista, com um sistema de avaliação externa com propósitos mais enquadráveis numa avaliação de cariz psicométrico. Tal como, aliás, é hoje referido por vários autores como uma inevitabilidade ou como uma situação que dificilmente terá uma solução de natureza muito diferente nos próximos anos (Gipps e Stobart, 2003; Kellaghan e Madaus, 2003; Torrance, 2003). E porquê? Fundamentalmente pelas questões relacionadas com a falta de fiabilidade, equidade e comparabilidade da avaliação interna, retirando-lhe, por isso, a credibilidade exigida pelas sociedades. Daí as pressões exercidas sobre os sistemas educativos para que mantenham e desenvolvam qualquer forma de avaliação externa cuja standardização, ou a sua natureza uniforme, permite garantir padrões aceitáveis de fiabilidade, de equidade e de comparabilidade, apesar dos problemas de validade que normalmente estão associados a provas de avaliação em larga escala. A «conciliação» entre uma avaliação interna essencialmente formativa, orientada para melhorar as aprendizagens, motivar os alunos, regular os processos de ensino e de aprendizagem e uma avaliação externa de natureza normativa e sumativa com funções de moderação, de selecção, de controlo e de avaliação do sistema, terá, como veremos, diferentes naturezas na legislação subsequente.

O Despacho n.º 162/ME/91 mal viu nascer a luz do dia pois foi revogado passados nove meses e 10 dias (!) e substituído por dois Despachos Normativos: o n.º 98-A/92, de 19 de Junho, referente ao sistema de avaliação da educação básica e o n.º 338/93, de 21 de Outubro, referente ao sistema de avaliação do ensino secundário.

No que se refere à avaliação interna, o novo Despacho 98-A/92 manteve, no essencial, os princípios e orientações constantes no Despacho anterior. Porventura, sublinhou mais enfaticamente a inspiração cognitivista e construtivista e a concepção fundamentalmente positiva da avaliação das aprendizagens. Talvez por

estas razões, a avaliação aferida deixou de ter quaisquer efeitos sobre a progressão dos alunos e, consequentemente, sobre a certificação. Ou seja, foi decidido que deixaria de haver qualquer tipo de avaliação externa que tivesse efeitos sobre as classificações dos alunos da educação básica, particularmente no final do 3.º ciclo. Dito de outra maneira, a avaliação na educação básica passou a ser da total e exclusiva responsabilidade dos professores e das escolas sem qualquer mecanismo de controlo ou de moderação por parte da administração quanto à execução do currículo e à avaliação das aprendizagens dos alunos. O poder político dava assim a prova máxima de confiança aos professores e às escolas, ao confiar-lhes totalmente o exclusivo da responsabilidade no desenvolvimento do currículo e na avaliação que ocorrem nas salas de aula. Talvez por isso, a medida tenha sido bem acolhida nos meios educativos e académicos, não se tendo questionado a fiabilidade e a equidade das avaliações; de facto, o Despacho não previa quaisquer elementos de moderação interna ou externa que, de certo modo, contribuíssem para que os professores avaliassem tendo em conta *standards* e critérios de avaliação comuns ou semelhantes. Provavelmente esta preocupação justificou uma iniciativa do Instituto de Inovação Educacional (IIE), com a colaboração da Escola Superior de Educação de Castelo Branco, que consistiu numa formação em avaliação das aprendizagens destinada a um conjunto de cerca de 420 professores, organizados em três grupos, oriundos de escolas de todas as Direcções Regionais de Educação. Era suposto que, a partir desta formação, se estabelecessem redes de professores e de escolas da educação básica tendo em vista articular e melhorar as práticas de avaliação e, sobretudo, sensibilizar e mobilizar os professores para a importância de se porem realmente em prática as orientações constantes no Despacho 98-A/92. Foi uma iniciativa bastante valorizada pelos professores participantes, conforme se refere em Lemos (1993), que acabou por não ter a sequência planeada devido a alterações então surgidas na equipa dirigente do Ministério da Educação. Apesar disso, no âmbito do IIE, e de acordo com as competências que lhe estavam atribuídas no domínio da avaliação, mencionadas no próprio Despacho 98-A, foi ainda possível produzir e distribuir por todas as escolas um *dossier* com um significativo conjunto de informação teórica e prática sobre avaliação, várias brochuras de apoio e estudos destinados a preparar a elaboração das provas de aferição (e.g., IIE, s/data, 1992a, 1992b; Oliveira, Pereira e Fernandes, 1993, 1994). Posteriormente, o Instituto de Inovação Educacional, que, entre 1990/1991 e 1994/1995, chegou a publicar e a promover a publicação de dezenas de estudos de avaliação, foi deixando de desenvolver trabalho significativo nesta área, acabando mesmo por ser extinto.

O Despacho 98-A/92, de 19 de Junho, acabou por fazer o caminho que o anterior não conseguiu fazer pois prevaleceu durante exactamente 9 anos e 3 dias! Foi revogado com a recente publicação do Despacho n.º 30/2001, de 22 de Junho, no qual se afirma que, com as necessárias adaptações e melhorias, no essencial se reproduzem os princípios e orientações do seu antecessor. Na verdade, o Despacho 98-A/92 foi considerado um normativo claramente progressista, inspirado nos resultados da psicologia cognitiva sobre a aprendizagem, defendendo essencialmente uma concepção formativa da avaliação e a total autonomia dos professores e das escolas em matéria de avaliação das aprendizagens. Na sequência da sua publicação e do papel que então foi desempenhado pelo IIE em matéria de avaliação, gerou-se no país uma interessante dinâmica em torno da questão da avaliação das aprendizagens que se consubstanciou na realização de múltiplas acções de formação, encontros de natureza diversa, trabalhos de investigação e trabalhos de reflexão e de interpretação do Despacho (Alaiz, 1993; Alves, 1997; Antunes, 1995; Benavente, Alaiz, Barbosa, Campos, Carvalho e Neves, 1995; Boavida, 1996; Campos, 1996; Conceição, 1993a, 1993b; Fernandes, 1992a, 1992b, 1992c, 1993a, 1993b; 1994a, 1994b, 1994c, 1995; Fernandes, Neves, Campos e Lalande, 1996; Gil, 1997; Lemos, 1992, 1993; Lemos, Neves, Campos, Conceição e Alaiz, 1992; Lobo, 1996; Neves, 1996). Por outro lado, em 1992/1993, a Universidade Católica Portuguesa iniciou o funcionamento do Mestrado em Ciências da Educação na especialidade de Avaliação em Educação. Em 1993/1994 é também criado o Mestrado em Ciências da Educação (Avaliação em Educação) na Faculdade de Psicologia e de Ciências da Educação da Universidade de Lisboa. Dir-se-ia que foi nos anos 90 do século XX que a avaliação entrou inequivocamente nas agendas da investigação, da formação, da educação e da política educativa.

Voltemos então à evolução (?) da avaliação aferida. Recordemos que, em cerca de nove meses, sob um mesmo governo, a natureza da avaliação aferida mudou substancialmente. De componente externa da avaliação com efeitos na progressão e certificação dos alunos e com objectivos de avaliar a consecução de objectivos e a qualidade do sistema, para componente externa da avaliação sem quaisquer efeitos na avaliação dos alunos, mas mantendo as preocupações com a avaliação da qualidade do sistema. Estávamos em 1992. Em Setembro de 1994, o Despacho Normativo n.º 644-A/94 fazia alguns aditamentos ao Despacho 98-A/92 no domínio da avaliação interna. Criavam-se as provas globais, da inteira responsabilidade das escolas, com um peso 1 contra peso 3 da classificação de frequência no final do 3.º período do 9.º ano de escolaridade. Assim, os alunos que em 1995/1996 e em 1996/1997 frequentassem, respectivamente, o 8.º e o 9.º anos de

escolaridade, teriam que fazer provas globais escritas como parte integrante da avaliação sumativa interna. Os alunos do 8.º ano faziam apenas a prova de Ciências Naturais. Relativamente às provas de aferição não houve lugar a quaisquer aditamentos. A avaliação e o desenvolvimento do currículo permanecia integralmente sob o controlo dos professores e das escolas.

Quanto ao Despacho 338/93, consagrou a avaliação externa através de exames no final do ensino secundário, com efeitos na classificação final dos alunos, na certificação e ainda no acesso ao ensino superior mas também através de provas de aferição a administrar sempre que se julgasse necessário. O facto dos exames nacionais do ensino secundário constituírem uma tarefa que mobiliza uma quantidade muito significativa de recursos e ainda o facto de nunca verdadeiramente se ter definido uma política para as provas de aferição podem explicar porque é que nunca mais se ouviu falar de concretização da avaliação aferida no ensino secundário.

Após uma aplicação piloto de provas de aferição de Matemática e de Língua Portuguesa, que teve lugar em 1999, o Despacho n.º 5437/2000, de 18 de Fevereiro, vem definir as disciplinas em que as provas se realizarão, os anos de escolaridade e respectivos anos de aplicação. Fica a perceber-se que serão aplicadas a todos os alunos e não a uma amostra probabilística. Fica a saber-se que serão aplicadas, progressivamente, a alunos dos 4.º, 6.º e 9.º anos de escolaridade, de acordo com uma calendarização que se estende até 2001/2002. Foi o Despacho que ditou verdadeiramente o início efectivo da aplicação das provas de aferição no nosso país. Cerca de oito anos após a sua criação! Foi também o Despacho que talvez tenha ido mais longe quanto à definição precisa dos propósitos da avaliação aferida e quanto à necessidade de se estabelecerem critérios de qualidade, de rigor e de controlo do nível das aprendizagens dos alunos do ensino básico. Vejamos um extracto do Despacho (itálico meu):

«A *qualidade*, o *rigor* e a *pertinência* da avaliação constituem elementos determinantes para se aferir do modo como se operam os desempenhos dos alunos, em articulação coerente com a configuração do currículo.

No ensino básico, a avaliação surge como elemento regulador do processo de ensino-aprendizagem, assegurando que *a transição entre ciclos de escolaridade e a obtenção do diploma deste nível de ensino possa corresponder a reais saberes e competências*. (...)

(...) *A avaliação aferida visa permitir o controlo dos níveis de desempenho dos alunos e a avaliação da eficácia do sistema*, através da devolução dos resultados às escolas para enriquecimento das aprendizagens, no âmbito do desenvolvimento dos respectivos projectos educativos.

(...) A avaliação aferida (...) destina-se a *medir o grau de cumprimento dos objectivos essenciais* (...) com o propósito de *contribuir para a tomada de decisões no sentido de melhorar a qualidade das aprendizagens e reforçar a confiança social no sistema educativo*.» (Despacho n.º 5437, 2000, p. 4613)

É talvez o Despacho mais clarificador no que se refere à avaliação aferida, mas parece persistir a necessidade de uma maior clarificação em matérias relacionadas com os efeitos e consequências dos resultados das provas, ao nível das políticas educativas, das práticas da administração, das práticas das escolas, do envolvimento dos pais e professores ou da formação de professores. Parece sentir-se ainda a falta de referência a outras funções da avaliação aferida, tais como a função moderadora das práticas de avaliação interna ou a função motivadora e indutora de novas práticas de ensino dos professores, ou seja a sua função formativa.

No fundo, em todos os despachos se sente a ausência de respostas ou orientações para questões como:

1. O que fazer com os resultados das provas de aferição?
2. Quem deve analisar os resultados?
3. Como organizar a discussão dos resultados ao nível local, regional e nacional?
4. Que consequências deveremos retirar dos resultados ao nível do currículo nacional? E ao nível das práticas de desenvolvimento curricular nas escolas e salas de aula?
5. Como estarão asseguradas a validade, a fiabilidade, a equidade e a comparabilidade das provas? Serão provas justas e adequadas para os alunos? A correcção é credível?
6. Que sistema de administração das provas parece mais adequado? Qual a melhor periodicidade? Haverá vantagens na administração das provas ao universo dos alunos?

7. As provas devem manter-se exclusivamente nas disciplinas de Matemática e de Língua Portuguesa? Não será importante realizarem-se noutras disciplinas como as Ciências, as Línguas, a Geografia, a História e outras?

Alguns meses após a publicação do Despacho n.º 5437, o Conselho Nacional de Educação (CNE), através do ponto 50 do seu parecer n.º 3/2000, de 14 de Julho, sobre a *Proposta de Reorganização do Ensino Básico*, pronunciava-se sobre as provas de aferição nos seguintes termos:

«Com as adequações que a experiência recolhida vier a recomendar (e sem necessidade aparente de virem a ser confundidas ou modificadas no sentido dos exames tradicionais), as provas de aferição poderão vir a constituir-se em fonte de indicadores valiosos para as estruturas centrais que dirigem e regulam a educação, para as escolas e para a sociedade em geral. Os resultados poderão ainda constituir um recurso de elevada valia para conduzir a escola a novas práticas pedagógicas, ao permitirem a identificação das virtudes e das deficiências dos processos actuais.» (CNE, 2000, p.13021)

Dada a natureza, a abrangência e a relevância do parecer, talvez se pudesse esperar um pouco mais acerca da avaliação aferida, uma matéria que suscita tantas interrogações, a tantos níveis (e.g., desenvolvimento do currículo nacional, impacto sobre o ensino e a avaliação nas escolas, avaliação da qualidade do sistema, concepção, desenvolvimento, administração e correcção das provas, fiabilidade, validade, equidade, comparabilidade e adequação das provas). Não faltando competência ao CNE neste e noutros assuntos educativos, não foi, com certeza, possível ir mais além do que conjugar um conjunto genérico, consensual e abstracto de recomendações que provavelmente não ajudarão muito o Ministério da Educação a tomar as decisões que, nesta matéria, há muito se impõem.

Regressemos, mais uma vez, ao percurso legislativo da avaliação aferida.

Cerca de dois anos após a publicação do Despacho 5437, surge o Despacho n.º 2205/2002, de 2 de Janeiro, que determina que as provas de aferição, em cada ciclo de escolaridade, serão aplicadas a todos os alunos em dois anos consecutivos, seguidos de um ano de pausa, no qual só serão aplicadas a uma amostra de alunos do ciclo em causa. Ou seja, como todos os alunos do 4.º ano realizaram provas em 2000 e 2001, em 2002 as provas serão aplicadas a uma amostra de alunos deste ano. O mesmo se passando para os outros ciclos, a partir da primeira aplicação. Trata-se de uma louvável tentativa de organizar a administração das provas que, no entanto, contém uma opção bastante discutível. Ou seja, para além de persistir a indefinição e a falta de clarificação sobre um conjunto de matérias que acima se exemplificaram, não se vislumbra o alcance de, no ano da pausa, se administrarem as provas a uma amostra da população em causa. Mas, afinal, não chegou a acontecer como despachado, para além de 2002, porque logo surgiu o Despacho n.º 474/2003, de 18 de Dezembro, que mais não fez do que determinar que as provas seriam sempre aplicadas a amostras de alunos de cada um dos ciclos. Nestes dois últimos despachos, os decisores focaram a sua atenção na questão do âmbito de administração das provas; ou seja, amostra *versus* universo. No entanto, há uma certa fragilidade na argumentação utilizada, o que indicia que continuava a não existir uma visão estratégica e integrada para a avaliação das aprendizagens e, em particular, para a avaliação aferida. E também ausência de clareza conceptual, como aliás demonstra o Despacho n.º 1911/2004, de 13 de Janeiro, que vem alterar o anterior, referindo que, afinal, em 2003/2004 as provas serão aplicadas a todos os alunos do 9.º ano, e não a uma amostra, porque é preciso «aproximar o tipo de provas realizadas pelos alunos do 9.º ano do modelo de avaliação externa prevista para o ano de 2005» (Despacho n.º 1991, 2004, p.1489). Trata-se de uma função nova das provas de aferição: preparar os alunos para o exame! É algo estranho e, além do mais, bastante mais caro do que outras soluções possíveis que poderiam ser postas em prática com o objectivo de preparar os alunos para um dado tipo de exame. Mas o que parece ser verdadeiramente mais sério é a confusão conceptual. Aparentemente, há o entendimento de que a natureza das provas de aferição, sem quaisquer efeitos na progressão dos alunos, é idêntica à das provas de exame, com efeitos na progressão escolar dos alunos. A meu ver, as suas naturezas devem ser substancialmente diferentes porque algumas das suas principais funções também são consideravelmente diferentes. Só para dar um exemplo, as provas de exame, que vão ter impacto na vida escolar dos alunos, têm, entre outras, uma função sumativa e uma função selectiva. As provas de aferição não têm estas funções. Por isso mesmo, muitos investigadores defendem que é no âmbito deste tipo de provas que se poderão apresentar tarefas de avaliação mais diversificadas e abertas que normalmente não é aconselhável incluir em provas de exame com efeitos no progresso escolar dos alunos. E porquê? Fundamentalmente por razões que se prendem com a fiabilidade das correcções (quanto mais abertas são as tarefas, mais complicado se torna garantir níveis aceitáveis de fiabilidade

entre correctores porque é mais difícil elaborar critérios de correcção que suscitem utilizações razoavelmente uniformes), com os processos de administração (em provas sem efeitos no progresso escolar dos alunos pode aceitar-se uma maior flexibilização nos procedimentos de administração aproximando-os dos contextos mais familiares aos alunos como por exemplo serem os seus professores a administrar as provas) ou com a consistência curricular (neste tipo de provas podem testar-se tópicos ou temas curriculares normalmente não «cobertos» nas provas de exame tradicionais, por exemplo tarefas que exijam algum tipo de experiência ou método experimental, alargando assim a sua incidência curricular).

Dada a natureza dos despachos posteriores e anteriores ao Despacho n.º 5437/2000, verifica-se que este, apesar de tudo, se mantém como o que mais avançou em termos de definição e de concretização da avaliação aferida. Aliás, todos os despachos posteriores o utilizam como referência e, até ver, ainda não foi revogado.

Como se viu anteriormente, as provas de aferição são instituídas em 1992. Em 2000, ou seja, passados cerca de oito anos, são aplicadas as primeiras provas de Matemática e de Língua Portuguesa a todos os alunos do 4.º ano de escolaridade. Assim, a administração das provas, sempre em Matemática e em Língua Portuguesa, decorreu ou vai decorrer da seguinte maneira:

1. Em 2000, a todos os alunos do 4.º ano de escolaridade;
2. Em 2001, a todos os alunos dos 4.º e 6.º anos de escolaridade;
3. Em 2002, a todos os alunos dos 6.º e 9.º anos de escolaridade e a uma amostra de alunos do 4.º ano;
4. Em 2003, a amostras de alunos dos 4.º, 6.º e 9.º anos de escolaridade;
5. Em 2004, a amostras dos alunos dos 4.º e 6.º anos e a todos os alunos do 9.º ano.

Dos quatro anos de administração de provas de aferição já decorridos foram produzidos, até ao dia em que redijo este texto, dois relatórios referentes às administrações de 2000 e de 2001 que foram distribuídos pelas escolas (Ministério da Educação, 2000; Ministério da Educação, 2002). Os dados referentes à administração de 2002 foram há dias divulgados através da comunicação social. Quanto aos dados de 2003 ainda não se possuem quaisquer informações. Desconheço, neste momento, se os dados de 2002 e de 2003 serão publicados sob a forma de relatórios. Temos aqui outro problema relacionado com os sinais que o poder político está a querer, ou não, dar à sociedade em geral, aos professores, aos alunos e aos seus pais ou encarregados de educação. Que objectivos pretendemos alcançar quando analisamos os dados, produzimos um relatório e o distribuímos pelos serviços do Ministério da Educação, pelas escolas e pela comunicação social? Será que a nossa função está terminada? Que consequências, que acções se empreendem perante os resultados? Será que as escolas estão a fazer alguma coisa com os resultados? Com as provas? Há alguém ou alguma instituição dependente ou independente do Ministério da Educação a reflectir sobre os dados ou a fazer alguma coisa no sentido de se cumprirem os grandes propósitos das provas de aferição? Por outro lado, que sinal estaremos a dar a todos os intervenientes se não sentimos necessidade de publicar atempadamente os resultados e a sua análise? Não me surpreenderia que muitos professores, alunos ou pais pensem que as provas de aferição não servem para nada. As pessoas poderão pensar que se o próprio Ministério da Educação que produz e aplica as provas não parece ter pressa em analisar os resultados é porque não as valoriza. Então porque havemos nós de as valorizar? Elas nem contam para as classificações. Dirão as pessoas. Ora isto é muito negativo. As provas devem ser valorizadas, os seus propósitos muito claros e deve haver ilações e acções a empreender por iniciativa do Ministério da Educação que, nestas matérias, deve dar o exemplo. Será que estaremos a prestar um bom serviço à avaliação do sistema educativo, à prestação de contas do serviço público de educação à sociedade, aos professores, aos alunos e aos demais intervenientes, se estamos dois anos para divulgar os resultados das provas? Será que estaremos a dar um sinal positivo acerca da grande relevância que estas provas devem ter?

Atente-se, por exemplo, no impacto do *National Assessment of Educational Progress* (NAEP) nos Estados Unidos da América. Trata-se de uma avaliação externa que também não tem quaisquer efeitos sobre a progressão ou sobre as classificações dos alunos e que está em vigor desde 1969. Presentemente é administrada de dois em dois anos, a amostras de estudantes dos 4.º, 8.º e 12.º anos com, respectivamente, 9, 13 e 17 anos de idade, numa variedade de disciplinas (Beaton, 1997; Jones, 2003; Lapointe, 1986; Sebring e Boruch, 1991). Vale a pena perceber como são utilizados os resultados obtidos, quer pelo poder político quer pela comunidade dos investigadores quer ainda pelos professores e pelas escolas. É um programa que ganhou uma grande credibilidade e é muito respeitado pelos principais intervenientes porque, desde muito cedo,

definiu bem os seus propósitos, estabeleceu finalidades claras, soube sempre promover o envolvimento de investigadores na análise dos resultados e sempre publicou e divulgou amplamente os resultados. Não cabe aqui elaborar muito mais sobre este programa de avaliação externa que está exaustivamente caracterizado nas referências acima indicadas.

As provas de aferição voltam a ser referidas no Decreto-Lei n.º 6/2001, de 18 de Janeiro, que «estabelece os princípios orientadores da organização e da gestão curricular do ensino básico» e onde se reafirma que não produzem efeitos na progressão escolar dos alunos e que são um instrumento de desenvolvimento do currículo nacional. O Despacho Normativo n.º 30/2001, de 22 de Junho, que «estabelece os princípios e os procedimentos a observar na avaliação das aprendizagens assim como os efeitos dessa avaliação», no que se refere aos alunos do ensino básico, não faz referência alguma à avaliação aferida.

Entretanto, o Decreto-Lei n.º 209/2002, de 17 de Outubro, vem alterar o artigo do Decreto-Lei n.º 6/2001 referente à avaliação sumativa. Fica determinado que esta passa a ter uma componente interna, da responsabilidade dos professores e das escolas e uma componente externa, da responsabilidade do Ministério da Educação, traduzida em exames nacionais, a realizar no 9.º ano de escolaridade nas disciplinas de Matemática e de Língua Portuguesa. O Decreto é omissivo quanto a todo um conjunto de questões decorrentes da introdução destes exames, nomeadamente quanto ao regime de progressão dos alunos, à natureza e âmbito das provas, escalas de classificação ou peso dos exames para efeitos de certificação. Devem ser aspectos a regulamentar posteriormente. Também nada refere sobre a avaliação aferida, particularmente no 9.º ano de escolaridade.

Posteriormente, a Lei n.º 31/2002, de 20 de Dezembro, aprova o sistema de avaliação da educação e do ensino não superior, desenvolvendo o regime previsto na Lei de Bases do Sistema Educativo. Julgo tratar-se de uma Lei que, quanto à forma e quanto ao conteúdo, não será muito feliz, pois não clarifica nem apresenta uma visão estratégica e integrada da avaliação da educação e do ensino não superior que nos continua a faltar. Cria, no âmbito do CNE, uma comissão especializada permanente para a avaliação do sistema educativo que deve ter sido constituída em Abril de 2003...

Ao fim de treze anos a avaliação aferida não tem o lugar credível e prestigiado que deveria ter no sistema educativo português. Penso não exagerar se afirmar que, para além daqueles que elaboram as provas, as administram e corrigem e dos que analisam os dados, ninguém verdadeiramente parece dar-lhes qualquer importância. Tenho tido a sensação que, do lado da administração, a avaliação aferida faz-se mais ou menos para cumprir um calendário que ainda nem acertado está. Não se tem feito a pedagogia necessária para se entender o alcance que estas provas podem ter na melhoria da qualidade da educação básica, na formação de professores, no desenvolvimento de bancos de itens a utilizar pelos professores e pelas escolas ou no desenvolvimento da investigação numa área em que estamos francamente atrasados em relação aos nossos parceiros da União. Só para termos uma noção do problema, uma investigação em curso que actualmente coordeno mostra-nos que, nos dez anos compreendidos entre 1993 e 2003, não foram elaboradas no nosso país mais do que 40 teses de mestrado na área da avaliação das aprendizagens. A avaliação aferida não foi objecto de qualquer investigação (C. Martins, comunicação pessoal, Fevereiro, 2004). Ora esta situação não favorece o nosso desenvolvimento pedagógico, científico, tecnológico e técnico numa área relevante da educação, abrindo caminho à pura especulação acerca dos processos e resultados do sistema educativo.

Se bem enquadrada, devidamente organizada e levada muito a sério, a avaliação aferida pode ser uma forma relativamente barata de se obter informação de elevada qualidade sobre vários desempenhos do sistema e uma alavanca importantíssima para, entre outros efeitos, induzir boas práticas de ensino e de avaliação, promover investigação ou apoiar a tomada de decisões políticas e administrativas aos níveis local, regional e nacional. Treze anos é muito tempo. O trabalho até agora realizado na avaliação aferida é mais do que modesto. A administração das provas só se iniciou em 2000, após terem sido criadas em 1992! Será que vamos prosseguir neste ritmo para, de uma vez por todas, definirmos claramente o papel da avaliação aferida no nosso sistema educativo?

Estudos Internacionais de Avaliação das Aprendizagens

Os estudos internacionais de avaliação, que comparam as aprendizagens adquiridas pelos alunos nos sistemas educativos, têm vindo a ganhar uma importância e uma notoriedade crescentes, particularmente a partir dos anos 80 do passado século. Há várias razões que podem explicar este facto, algumas das quais já acima foram aduzidas a propósito da necessidade de mudar a avaliação.

As sociedades de hoje são mais interdependentes dos pontos de vista económico, social, político e tecnológico, são mais abertas e as pessoas têm uma capacidade muito maior de mobilidade o que gerou movimentos migratórios importantes, particularmente de países da Europa de Leste, de países Latino-Americanos, de países Africanos e mesmo de países Asiáticos para os países mais desenvolvidos do chamado Mundo Ocidental. Como consequência destes fenómenos, as sociedades de hoje são multiculturais, mais competitivas e mais exigentes a todos os níveis. Os sistemas educativos confrontam-se com populações de estudantes que nada têm a ver com as populações, essencialmente homogêneas, de há 30 ou 40 anos atrás.

Muito provavelmente este incremento de estudos de avaliação terá um pouco a ver com tudo isto, pois os países precisam que os seus sistemas educativos respondam com qualidade às necessidades de formação, às legítimas aspirações e às motivações dos jovens. Porque, desta forma, estarão também a responder aos desafios da competitividade, da mobilidade e da interdependência a todos os níveis.

Há ainda outras razões, eventualmente decorrentes das anteriores, para o que parece ser o reforço recente destes estudos. Por exemplo, os decisores políticos começaram a interessar-se pela avaliação, pressionados pelas exigências do mundo da economia, pelos meios de comunicação social e pelos próprios cidadãos que, cada vez mais, querem saber se os serviços públicos de educação, de saúde ou de justiça, prestam o serviço que são supostos prestar. Os estudos internacionais podem pressionar os mais diversos agentes do sistema educativo no sentido de alterarem as suas práticas e procedimentos, podem «ajudar» os governos a justificar ou a fundamentar mudanças nos currículos, nos sistemas de formação de professores, nos sistemas de alocação de recursos às escolas ou nos sistemas de gestão e administração escolar. Mas podem também constituir uma pressão sobre os próprios governos para que adequem e alterem as suas políticas educativas.

Há duas organizações internacionais de natureza bem diferente que se têm destacado na realização de tais estudos. Uma, a *International Association for the Evaluation of Educational Achievement* (IEA), é uma organização não governamental que foi fundada em 1958, após uma reunião de investigadores e especialistas em assuntos de educação, que teve lugar no Instituto de Educação da UNESCO, em Hamburgo. Presentemente, integram a IEA 55 países por via dos respectivos Ministérios da Educação ou de institutos de investigação mais ou menos independentes. De acordo com Plomp, Howie e McGaw (2003) a sua grande finalidade é a de, através dos estudos que promove, contribuir para melhorar a qualidade da educação. Os dois propósitos dos seus estudos internacionais são os seguintes:

1. Proporcionar informação aos decisores políticos e à comunidade educativa acerca da qualidade da educação, tendo por referência países que possam ser relevantes para o efeito;
2. Apoiar os países participantes a compreender as diferenças observadas entre si e as que se verificam dentro de cada país.

Em conformidade com estes propósitos, a IEA faz dois tipos de comparações: num caso, compara internacionalmente os efeitos da educação a que as crianças e os jovens têm acesso nos diferentes países, com base nos resultados que os alunos obtêm em testes internacionais; no outro, analisa até que ponto, em cada um dos países, o currículo proposto para um dado nível de escolaridade é realmente posto em prática nas escolas e é realmente aprendido pelos alunos.

A IEA começou a promover estudos internacionais em 1959 com a preocupação de comparar os resultados dos sistemas educativos como um todo e não os resultados dos estudantes ou das escolas tomados individualmente. O primeiro estudo internacional de avaliação das aprendizagens na disciplina de Matemática foi realizado pela IEA nos anos 60. Até aos anos 80, os estudos eram essencialmente controlados pelos interesses dos investigadores nas áreas do currículo, da construção de testes e da análise e interpretação de dados. A partir daquela década começam a surgir também os interesses dos decisores políticos e de diferentes grupos e interesses sociais, pelas razões que acima se enumeraram.

De acordo com Plomp et al (2003), as quatro questões fundamentais do *Third International Mathematics and Science Study* (TIMSS), o último estudo realizado pela IEA, na segunda metade dos anos 90, destinado a

alunos de 9 anos (3.º ou 4.º anos de escolaridade), de 13 anos (7.º ou 8.º anos de escolaridade) ou no último ano do ensino secundário, eram as seguintes:

1. O que é que se espera que os alunos aprendam? (Ou seja, o que é que está previsto no currículo? Que conteúdos? Que finalidades? Que recomendações?)
2. Quem é que assegura o processo de ensino? (Ou seja, quem são os professores? Que formação possuem? Que concepções e saberes revelam?)
3. Como é que se organiza o ensino? (Ou seja, que métodos e estratégias são utilizadas? Que dinâmicas de sala de aula se desenvolvem? Qual o papel do professor e o papel dos alunos?)
4. O que é que os alunos aprendem efectivamente? (Ou seja, o que é que os alunos aprendem ao fim de um dado período de tempo relativamente ao que está previsto no currículo?)

Repare-se que a natureza das questões revela uma preocupação em analisar os currículos dos países participantes, as formas encontradas para os pôr em prática, os contextos de implementação e, naturalmente, as aprendizagens adquiridas pelos alunos ao fim de um certo período de ensino. Em geral, todos os estudos da IEA tinham estas características genéricas, sendo certo que no TIMSS se foi mais longe na tentativa de ter mais em conta os contextos em que o ensino se desenvolve nos diferentes países e, por isso, foi estudado um conjunto de variáveis relacionadas com as escolas, os professores e os alunos.

Para o efeito deste texto o que interessa talvez sublinhar é que os estudos da IEA são baseados nos currículos em vigor nos países participantes, que são analisados ao nível do sistema (currículo proposto), ao nível da escola ou da sala de aula (currículo ensinado ou currículo posto em prática) e ao nível do aluno (currículo aprendido), tendo em conta os diferentes contextos que podem influenciar o desenvolvimento do currículo, tais como as condições das escolas e das salas de aula, os processos utilizados, as características das comunidades em que as escolas se inserem assim como as características das escolas, dos professores e dos alunos.

Trata-se de um esforço interessante se pensarmos que, nos estudos da IEA, participam dezenas de países diferentes. Não irei entrar aqui em pormenores técnicos, mas poderei dizer que tem havido uma evolução sensível nos mecanismos de controlo relativamente a questões mais ou menos complexas, como é o caso do processo de amostragem, da decisão acerca das populações que devem participar no estudo (a questão da idade *versus* ano escolaridade), da selecção dos itens, da tradução dos itens ou das garantias de que todos os procedimentos são seguidos pelos diferentes países da mesma forma.

A outra organização que tem promovido estudos internacionais de avaliação das aprendizagens dos alunos é a Organização para a Cooperação e Desenvolvimento Económico (OCDE) que integra 30 países membros, da Europa, da América do Norte e da Ásia (Pacífico) que partilham as ideias do mercado livre e da democracia pluralista (OECD, 2000). A abordagem da OCDE às questões da educação e, em particular, da avaliação das aprendizagens proporcionadas pelos sistemas educativos, começou por se centrar na recolha e tratamento de dados fornecidos pelos países membros.

Segundo Kellaghan (1986), na sequência de um conjunto de preocupações com a qualidade da educação manifestadas pelos ministros da educação dos países da OCDE, em 1984, seguiu-se um interesse crescente por parte da organização pelos estudos comparativos internacionais. A partir de 1992, começou a publicar anualmente *Education at a Glance* onde, em cada número, se faz uma comparação internacional, baseada num conjunto de dezenas de indicadores, que proporcione uma visão acerca do funcionamento dos sistemas educativos, nomeadamente através da análise dos recursos investidos na educação e do tipo de resultados daí decorrentes. Por exemplo, em OECD (1997; 2000), podem identificar-se algumas dezenas de indicadores no conjunto das seguintes áreas:

1. Contexto em que os sistemas educativos funcionam;
2. Recursos humanos e financeiros investidos na educação;
3. Acesso à educação, participação, progressão e conclusão;
4. Ambiente de aprendizagem e formas de organização dos sistemas educativos;
5. Resultados da educação ao nível individual, social e do mercado de trabalho;
6. Aprendizagens adquiridas pelos alunos.

O exaustivo conjunto de indicadores resultantes destas áreas permite que os países se comparem entre si e que, supostamente, possam fundamentar medidas de política que considerem apropriadas à situação ou aos problemas detectados. Tratava-se de um trabalho de natureza mais geral do que os estudos da IEA mas, sob muitos pontos de vista, com indicações de tendências e informações consideradas úteis por parte dos países membros.

Na segunda metade dos anos 90 a OCDE decide avançar com estudos internacionais relativos à avaliação das aprendizagens adquiridas pelos alunos, através do *Programme for International Student Assessment* (PISA). Na verdade, os resultados dos alunos utilizados em edições do *Education at a Glance* eram recolhidos dos estudos desenvolvidos no âmbito da IEA, como é o caso do TIMSS no final da década de 90, ou o caso do *International Assessment of Educational Progress* (IAEP), em que Portugal também participou nos meados dos anos 80. Talvez por esse facto a OCDE entendeu que deveria desenvolver estudos que fossem mais consistentes com os tipos de preocupações e o tipo de cultura da organização, mais ligada ao mundo do trabalho, com claras preocupações económicas e com o retorno dos vultuosos investimentos em recursos financeiros e em recursos humanos que os países fazem na educação. A cultura da IEA é mais marcada pelos investigadores e académicos da educação e daí talvez o enquadramento conceptual dos seus estudos ser baseado na teoria curricular.

A abordagem da OCDE é diferente da abordagem da IEA. O seu enquadramento conceptual, se assim se pode dizer, está baseado no conjunto de indicadores que se foram produzindo na organização, nomeadamente para o *Education at a Glance*. Com efeito, como referido em Plomp, Howie e McGaw (2003), um conjunto de 38 indicadores eram organizados em três categorias gerais para o caso do *Education at a Glance*:

1. Contextos da Educação

- a) Contexto demográfico (3 indicadores)
- b) Contexto social e económico (4 indicadores)

2. Custos, Recursos e Processos Escolares

- a) Despesas com a educação (8 indicadores)
- b) Recursos humanos (2 indicadores)
- c) Participação na educação (7 indicadores)
- d) Características da tomada de decisões (4 indicadores)

3. Resultados da Educação

- a) Resultados dos alunos (4 indicadores)
- b) Resultados do sistema (4 indicadores)
- c) Resultados do mercado de trabalho (2 indicadores)

Mais tarde, e ainda de acordo com Plomp et al (2003), o Comité de Educação da OCDE decidiu, em 1999, adoptar um enquadramento baseado em cinco grandes áreas:

- 1. Contextos que possam ser relevantes para os resultados;
- 2. Investimentos em recursos humanos e financeiros;
- 3. Acesso à educação e à aprendizagem; participação, progressão e conclusão;
- 4. Ambiente de aprendizagem e organização das escolas;
- 5. Resultados da educação ao nível individual, social e do mercado de trabalho e transição da escola para o trabalho.

Plomp et al (2003) referem o facto de, por exemplo, nenhum destes enquadramentos constituir um modelo conceptual que mostre como é que estão ligadas componentes dos sistemas educativos, tais como os *inputs*, os processos e os resultados.

O PISA está naturalmente influenciado por este trabalho anterior da OCDE e, como se disse, é substancialmente diferente do TIMSS. Pretende averiguar em que medida os jovens de 15 anos de cada país estão preparados para enfrentar os desafios das sociedades dos nossos dias. Por isso, a abordagem de avaliação do estudo não é baseada no que os alunos sabem acerca dos conteúdos constantes nos currículos, como era o caso do TIMSS, mas, no dizer da OCDE, vai para além de uma abordagem escolar, procurando avaliar a utilização que os alunos fazem do conhecimento em tarefas do dia-a-dia (OECD, 2003).

A recolha de dados do PISA ocorre em três ciclos separados por três anos. A primeira recolha ocorreu em 2000, a segunda em 2003 e a terceira ocorrerá em 2006. Em cada um destes ciclos recolhem-se dados sobre as competências dos alunos em Matemática, Ciências e Literacia. No primeiro ciclo, a ênfase foi na Literacia o que significa que dois terços dos testes incidiram nesta área e um terço nas áreas da Matemática e das Ciências. No segundo ciclo, que está a decorrer, a ênfase é na Matemática. Refira-se que neste ciclo ocorrerá uma recolha de dados acerca das competências desenvolvidas pelos alunos na resolução de problemas, domínio transversal.

As avaliações do PISA permitem obter três tipos de resultados:

1. **Indicadores básicos**, que proporcionam um perfil do conhecimento e competências dos alunos;
2. **Indicadores de contexto**, que mostram como tais competências estão relacionadas com importantes variáveis demográficas, sociais, económicas e educacionais;
3. **Indicadores de tendências**, que resultam da natureza continuada da recolha de dados e que mostram alterações nos resultados e nas distribuições e nas relações entre variáveis e resultados ao nível dos alunos e ao nível das escolas. (OECD, 2003, p.13)

Reconhecendo que os indicadores não são normalmente capazes de responder a questões prementes de política educativa, a OCDE/PISA desenvolveu um plano de análise orientado para a decisão política que irá para além da apresentação de indicadores. (OECD, 2003)

Apesar das suas diferenças quanto a vários aspectos relativos à concepção e desenvolvimento dos estudos, quer a IEA quer a OCDE visam comparar as aprendizagens adquiridas pelos alunos em diferentes países, caracterizar quais são os factores que estão mais relacionados com o desenvolvimento das aprendizagens de país para país e verificar se, em todos eles, são os mesmos factores ou são factores diferentes.

Ao nível dos impactos e das utilizações destes estudos nas práticas e nas decisões políticas dos países participantes, Kellaghan (1996) e Plomp et al (2003) referem, por exemplo, os seguintes:

1. Permitir a identificação de aspectos dos sistemas educativos que podem ser considerados problemáticos porque são discrepantes em relação ao que se passa na maioria dos países (e.g. conteúdos curriculares, níveis das aprendizagens adquiridas pelos alunos);
2. Monitorizar o desenvolvimento dos sistemas através de um ciclo de avaliações regulares nas disciplinas que estão a ser objecto de avaliação, para que possa haver séries de dados que indiquem as tendências de evolução dos sistemas. Trata-se do que neste momento se está a passar com os ciclos de estudos da IEA e da OCDE em Matemática, em Ciências e na Literacia.
3. Compreender diferenças entre os sistemas educativos e as diferenças dentro de cada um deles que podem contribuir para a tomada de decisões acerca da organização da escolaridade, da gestão dos recursos disponíveis ou sobre a organização das práticas de ensino.
4. Contribuir para desenvolver uma cultura de avaliação e para o esclarecimento fundamentado acerca do desenvolvimento dos sistemas através da difusão de ideias junto das organizações que tomam decisões. Significa isto que podem contribuir para enriquecer a discussão pública acerca da educação. Os estudos podem ainda contribuir para que os decisores percebam mais claramente o que, por exemplo, as escolas procuram alcançar, o que efectivamente alcançam e o que lhes será possível alcançar.
5. Desenvolver capacidade de investigação e capacidades técnicas e tecnológicas em países com menos competências em áreas como os processos de amostragem, metodologias de avaliação em larga escala, processos de correcção, tratamento estatístico dos dados, análise dos dados ou identificação de populações.
6. Proporcionar a países que não possuíam quaisquer dados sobre as aprendizagens desenvolvidas pelo sistema, a possibilidade de os começarem a ter e de os poderem passar a obter de forma mais regular, através das competências que, entretanto, foi possível adquirir.
7. Tornar os países mais informados e conscientes do que se passa à sua volta e retirar daí as necessárias ilações.

Keeves (1995), a propósito dos impactos dos estudos internacionais, refere que um significativo número de países, tais como a Austrália, a Hungria, a Irlanda, o Japão, a Nova Zelândia e os Estados Unidos, procederam a mudanças curriculares, mais ou menos profundas, na sequência da sua participação em estudos da IEA. Refere ainda que, uma vez que a decisão política não é normalmente documentada nem publicada, é natural que possam existir outros efeitos que não são propriamente publicitados. Por outro lado, os estudos podem não responder directamente às questões que os decisores gostariam de ver respondidas, mas acabam por ajudar a fundamentar o planeamento e a tomada de decisões.

Também Robitaille, Beaton e Plomp (2000) se referem ao impacto do TIMSS sobre o ensino e a aprendizagem da Matemática e das Ciências, mencionando a quantidade e a natureza das muitas discussões que tiveram lugar em muitos países participantes.

Portugal, como se sabe, tem participado e está neste momento a participar, quer no TIMSS quer no PISA, agora na sua segunda fase. Participou ainda no *International Assessment of Educational Progress*, nos finais dos anos 80. Tal como em relação às provas aferidas, fica-se com a sensação de que a nossa participação neste tipo de estudos, apesar das suas limitações, como mais abaixo terei oportunidade de identificar, tem sido algo inconsequente. Isto é, não tem obedecido a qualquer plano deliberado e estratégico que permita atingir objectivos relevantes para o nosso sistema educativo. Na verdade, para além de todos aqueles que estão directamente envolvidos na concretização dos estudos, poucos são os que sabem o que quer que seja sobre a sua natureza, os seus objectivos, as suas questões de investigação, as suas metodologias ou sobre as suas limitações. Como resultado deste tipo de situação abrem-se as portas para considerações e apreciações sobre a nossa participação, que ignoram aspectos relevantes da análise e da interpretação dos dados.

Por exemplo, na primeira série do PISA, Portugal aparece no *ranking* de países numa posição modesta em Matemática. Como já se referiu, o PISA testa jovens de 15 anos. Acontece que, na maioria dos países participantes, os jovens desta idade frequentam o 10.º ano de escolaridade ou, no mínimo, o 9.º. Em Portugal, temos um número muito elevado de jovens de 15 anos que frequentam os 7.º, 8.º ou 9.º anos de escolaridade ou mesmo anos de escolaridade mais iniciais. Muitos deles foram testados e, naturalmente, os resultados foram fortemente afectados pois é muito diferente testar um jovem de 15 anos que frequenta o 8.º ano de escolaridade ou o 10.º ano de escolaridade. Uma análise mais fina dos dados mostra que a média dos jovens portugueses que frequentam o 10.º ano de escolaridade é superior à média dos países participantes. O que pode ser também interessante analisar e interpretar. Em todo o caso, o que aqui se pretende referir é a incapacidade que, ao longo dos anos, temos vindo a demonstrar para proceder a análises nacionais dos dados. Não para esconder os dados internacionais, mas para contextualizar a análise e proceder a análises que retratem mais rigorosamente a nossa realidade e que, no contexto de um estudo internacional, poderão não fazer muito sentido. Por outro lado, persiste a nossa incapacidade de promover discussões de diversas naturezas acerca da participação nestes estudos: com investigadores, com professores, com educadores, com pais, com *opinion-makers* ou com quaisquer grupos sociais ou profissionais. É de certo modo confrangedor ver os resultados dos estudos serem sempre divulgados sob a forma de quadros do tipo das classificações das equipas de futebol a partir dos relatórios internacionais, quando a complexidade, a abrangência e a profundidade dos estudos deveria ser abordada de outra forma.

Ou seja, Portugal participa neste tipo de estudos desde a década de 80. Há cerca de 20 anos e, até hoje, não se conhecem consequências significativas e visíveis de tais participações. Eventualmente existirão algumas mas, em geral, o panorama não me parece positivo. É preciso pensar se estamos nestes estudos para cumprir calendário, porque «não podemos deixar de estar», porque «parece mal não estarmos» ou porque queremos que eles constituam mais uma oportunidade para conhecer melhor e desenvolver o sistema educativo.

Apesar do cuidado com que, hoje em dia, estes estudos são desenvolvidos e apesar da grande evolução que, nestes últimos 40 anos, se tem verificado ao nível da construção de itens e de testes, da administração das provas, da elaboração e utilização dos critérios de correcção, da análise de resultados ou da contextualização das avaliações, persistem problemas que ainda não estão resolvidos e que têm que ser tidos em conta.

Beaton, Postlethwaite, Ross, Spearritt e Wolf (2000), Goldstein (1996), Kellaghan (2003) e Kellaghan e Grisay (1995) referem problemas que persistem nos estudos internacionais de avaliação das aprendizagens adquiridas pelos alunos. Indicam-se aqui alguns deles:

1. Sendo os sistemas educativos dos vários países diferentes quanto aos objectivos que definem, às ênfases curriculares, aos tempos e espaços que atribuem aos diferentes domínios do currículo, torna-se muito difícil desenvolver uma metodologia que possa avaliar adequadamente os resultados em tal diversidade de currículos e contextos;

2. Tendo em conta as diferenças estruturais e organizacionais dos diversos sistemas educativos e ainda as diferenças culturais e sociais entre os países, não é fácil garantir a equivalência das populações e das amostras de estudantes que se pretendem comparar. Por exemplo, pensemos nos países em que a escolaridade formal se inicia aos 7 anos de idade e nos que se inicia aos 6 anos de idade. Ou pensemos no que se espera socialmente de um jovem de 15 anos nuns países e noutros.
3. Continua a ser difícil garantir que a tradução dos instrumentos tenha o mesmo significado ou seja interpretada da mesma forma em todos os países.
4. Sendo os resultados apresentados em termos de uma única classificação numérica que se refere às aprendizagens, há problemas quanto aos pressupostos relativos à natureza da aprendizagem.
5. Não está garantido que os itens que integram os testes tenham o mesmo tipo de relevância curricular nos diferentes países. Uns, podem dar-lhe uma grande importância e, outros, podem dar-lhe pouca ou mesmo nenhuma.
6. Vários países compararam os resultados dos seus alunos em diferentes estudos internacionais e verificaram que não havia consistência. Os investigadores concluíram que parecia improvável que tais discrepâncias reflectissem verdadeiras diferenças nas aprendizagens, inclinando-se para variações dos estudos, quanto aos conhecimentos e processos avaliados, correcção das respostas ou aos critérios utilizados na definição dos níveis de proficiência.
7. Parece não estar ainda garantido que todos os países participantes sigam fielmente os procedimentos emanados da coordenação dos estudos. Particularmente no que se refere aos procedimentos de identificação da população e respectiva selecção da amostra, de administração e correcções dos testes ou de tradução.

Todos estes problemas têm que ser considerados numa análise isenta dos resultados dos estudos internacionais. Parece-me que, do mesmo modo que não devemos ignorar os seus resultados, a qualidade de muitos dos seus procedimentos e o saber, a competência e a integridade dos seus investigadores responsáveis, também não devemos ignorar as suas limitações. Algumas das quais podem mesmo invalidar quaisquer inferências minimamente credíveis acerca das aprendizagens dos alunos.

Em Portugal, tanto quanto posso perceber, não se têm considerado ou discutido devidamente estas limitações. A ideia que se tem é que o nosso país tem problemas de diversa ordem (por exemplo, alunos de uma dada idade distribuídos por vários anos de escolaridade ou ausência de qualquer prática na resolução de testes de escolha múltipla da grande maioria dos alunos da educação básica) que o poderão colocar, à partida, numa posição desvantajosa. Valia a pena estudar ponderadamente estes e outros problemas decorrentes da nossa participação nos estudos. Na verdade, quando os resultados destes estudos são divulgados, normalmente de forma muito negativa e até alarmista, surgem invariavelmente duas posições que, a meu ver, em nada contribuem para a discussão fundamentada e clarificadora que se impõe. Uma, é a de rejeição pura e simples dos resultados dos estudos, com base numa ou mais das suas conhecidas limitações. Outra, é a de os aceitar como indicadores indesmentíveis e absolutos dos falhanços da escola «dominada pelos excessos das ciências da educação», utilizando-os, assim, como arma de arremesso político contra os que sustentam uma concepção do currículo e da escola que desenvolva, em **todos** os alunos, um alargado espectro de aprendizagens que lhes permitam integrar-se plenamente na sociedade. Ora, nem uma nem outra destas posições habituais contribui para que se analisem os resultados com a profundidade, a imparcialidade e a objectividade aconselháveis. E, muito menos, para que delas se retirem as consequências que se revelem consistentes com uma visão estratégica de modernização, de democratização e de desenvolvimento dos sistema educativo português. Parece-me que vai sendo tempo para que a discussão sobre estas questões tenha lugar noutro patamar de qualidade. O Ministério da Educação tem aqui um importante papel a desempenhar porque lhe compete definir uma política clara sobre a participação do nosso país neste tipo de estudos.

UMA AGENDA, MUITOS DESAFIOS

O presente texto aborda um conjunto relativamente alargado de questões de natureza teórica e prática que se referem à avaliação das aprendizagens. Houve desde o início dois tipos de preocupações. Por um lado, definir «coordenadas» teóricas que fundamentem discussões, afirmações e pontos de vista. Por isso, surge uma secção com referências relativamente recentes, que me pareceram relevantes, e também alguma discussão em torno dos paradigmas que «sustentam» a avaliação psicométrica e a avaliação alternativa. Por outro lado, abordar questões de natureza prática, discutindo princípios que devem orientar a avaliação alternativa, integrada no ensino e na aprendizagem, da responsabilidade dos professores e das escolas e analisando práticas de avaliação de professores.

Em cada uma das áreas da avaliação das aprendizagens o texto identifica problemas que se discutem na literatura, práticas que se têm desenvolvido e questões que se mantêm em aberto. Pode dizer-se que este texto teve por objectivo enumerar um alargado, mas não exaustivo, conjunto de matérias que, a meu ver, convém analisar e discutir de forma mais sistemática, mais aprofundada e mais aberta.

Há questões que aqui não foram discutidas, algumas das quais a necessitar que isso se faça com alguma urgência e em profundidade, como é o caso das características psicométricas dos exames, dos efeitos dos exames nas escolas e nas salas de aula ou da consistência das avaliações realizadas nas salas de aula. Há muito por fazer e a fazer numa área que parece algo desagregada, algo ao sabor de circunstâncias, sem rumo bem definido.

Por outro lado, o texto, ao enumerar e discutir um conjunto de assuntos relacionados com a avaliação das aprendizagens, tinha por objectivo definir uma agenda de acção que identificasse e sintetizasse a natureza dos problemas e que, simultaneamente, contribuisse para traçar caminhos de desenvolvimento numa área que, porque não dizê-lo, tem sido encarada com algum desleixo.

A agenda acaba, inevitavelmente, por lançar desafios às investigadoras e investigadores desta área e às instituições de ensino superior, às escolas e às suas professoras e professores e também aos responsáveis políticos e técnico-pedagógicos do Ministério da Educação.

Uma das primeiras constatações que me parece poder ser feita com legitimidade é a ausência quase total de investigação consequente no domínio da avaliação das aprendizagens. Há algumas teses de mestrado em curso, outras realizadas na última década do passado século e um ou outro projecto de investigação, mas não há quaisquer linhas ou programas de investigação que obedeçam a qualquer plano que estabeleça prioridades, que defina objectivos e que esteja articulado com reais necessidades de desenvolvimento do sistema educativo nesta área. O que há é uma dispersão exagerada de pequenos projectos e de teses de mestrado mais ou menos desgarradas umas das outras. Como consequência, há muito poucas publicações na área da avaliação das aprendizagens que sejam baseadas em dados empíricos. A maioria são reflexões teórico-práticas, por vezes baseadas em revisões de literatura mais ou menos profundas.

Nestas condições, parece prioritária a definição de linhas de trabalho investigativo que se centrem nas salas de aula e nas escolas. Precisamos de conhecer melhor o que é que os professores fazem no domínio da avaliação das aprendizagens. Precisamos de investigar questões, tais como:

1. Como é que os professores integram a avaliação no ciclo do ensino e da aprendizagem?
2. Que critérios de avaliação utilizam e como os articulam com o desenvolvimento das aprendizagens estruturantes que os alunos têm que realizar?
3. Que estratégias, técnicas e instrumentos de avaliação privilegiam?
4. Qual o papel dos alunos, dos pais ou encarregados de educação e de outros professores no processo de avaliação?
5. Como se poderão caracterizar as tarefas de avaliação utilizadas pelos professores? Como são seleccionadas?
6. Há articulação entre os professores de uma escola quanto às suas práticas de avaliação e de ensino?
7. Que efeitos tem a avaliação externa nas práticas de ensino e de avaliação dos professores?
8. Os professores analisam os resultados dos seus alunos ou das suas escolas nas provas de avaliação externa? Utilizam-nos? De que formas? Que articulações estabelecem entre a avaliação interna e a avaliação externa?

Estas e outras questões devem permitir a identificação de linhas de investigação que se reflectam na organização e funcionamento de programas de pós-graduação já criados ou a criar e que contribuam para uma melhor definição do trabalho investigativo.

A análise das poucas investigações em avaliação das aprendizagens que se fizeram em Portugal, os dados das provas de aferição do ensino básico e dos exames do ensino secundário sugerem que é necessário um esforço sério de formação. Mas não uma formação generalista sobre avaliação! A questão está em contextualizar, em concretizar, nas diferentes disciplinas, estratégias de avaliação que integrem o ensino e a avaliação e que respondam às dificuldades sentidas por muitos professores em promoverem nas suas aulas uma avaliação alternativa mais válida, mais abrangente, mais transparente, mais exigente, mais diversificada e com tarefas mais ricas dos pontos de vista educativo e formativo. A avaliação tem como uma das suas funções ajudar os alunos a aprender. Logo, é desejável que qualquer formação em avaliação integre equipas de formadores do currículo, da avaliação e das didácticas das disciplinas.

Por outro lado, a formação tem que estar bem relacionada com a investigação e, naturalmente, com as práticas. Os processos de formação deveriam ter um enquadramento tal que permitisse abordagens alternativas que nada têm a ver com «turmas de professores» a ouvir o que os formadores têm para dizer. Ou seja, abordagens em que a formação é feita **com** os professores e não **para** os professores, num processo em que as práticas não podem deixar de ser um elemento que contextualiza e dá real significado a todo o conjunto de perspectivas teóricas e a todo o conjunto de discussões e reflexões que a formação deve proporcionar. A formação só tem real sentido se estiver devidamente articulada com os processos de investigação. Na verdade, é a partir da investigação que se pode sistematizar um importante conjunto de práticas, de saberes, de estratégias e de atitudes que ajudem a reconstruir concepções e práticas nos processos de formação. Por outro lado, a investigação, utilizando a formação como contexto, permite-nos perceber os significados que os professores atribuem a todo o conjunto de problemas que a avaliação das aprendizagens lhes coloca na organização do seu ensino. Em suma, há uma relação Formação-Investigação-Práticas que tem que ser aprofundada e devidamente tida em conta.

As escolas, através dos seus órgãos próprios, devem definir «políticas» de avaliação das aprendizagens que respondam às orientações do currículo nacional e aos seus projectos educativos. Tais «políticas» devem contribuir para a definição de critérios de avaliação devidamente articulados com as aprendizagens estruturantes e essenciais a desenvolver e com as tarefas que se devem propor aos alunos. Desta forma, as escolas podem contribuir para melhorar substancialmente a consistência das suas avaliações e, conseqüentemente, melhorar a sua credibilidade junto da comunidade educativa e da sociedade em geral. Há sistemas de moderação que a escola e os seus professores podem utilizar para que, pelo menos no que diz respeito às aprendizagens de conceitos, processos ou procedimentos estruturantes, haja alguma uniformização de procedimentos ao nível do ensino e da avaliação. A ideia é a de conseguir que os alunos tenham oportunidades iguais, ou muito semelhantes, para aprenderem o essencial e, nestas condições, garantir que a equidade da avaliação seja uma realidade.

Por outro lado, no que à avaliação externa diz respeito, as escolas devem prever mecanismos de análise e discussão dos seus resultados e dos resultados dos seus alunos para que, a partir daí, possam retirar as devidas consequências para a reformulação das suas «políticas» de ensino e de avaliação. As escolas não devem ficar indiferentes aos resultados das avaliações externas, quaisquer que eles sejam, pois, de certo modo, é uma forma de se verem ao espelho e decidirem se têm que fazer alguma coisa acerca da imagem que lhes é reflectida. Além disso, é importante que comparem o currículo que é avaliado através da avaliação externa com o que é avaliado através das suas avaliações internas, que analisem os efeitos das avaliações externas nas «políticas» da escola e, sobretudo, nos professores e no seu ensino, nos alunos e nas suas aprendizagens.

Naturalmente que há entidades do Ministério da Educação que não podem ficar indiferentes a estes desafios que as escolas têm que enfrentar para que contribuam efectivamente para a modernização de procedimentos, para que a equidade seja uma realidade e para que o ensino e as aprendizagens entrem noutra patamar de exigência e de qualidade.

O Ministério da Educação, como veremos noutros pontos desta agenda, não pode deixar as escolas entregues a si próprias em matérias com a relevância do ensino, da aprendizagem e da avaliação, a essência do processo de educação e de formação. Tem que produzir e distribuir materiais de qualidade nos domínios do ensino e da avaliação, como aconteceu no início dos anos 90 com a publicação do *dossier Pensar Avaliação, Melhorar a Aprendizagem*, no âmbito do extinto Instituto de Inovação Educacional. Mas é essencial que esses materiais sejam associados a processos de formação que envolvam activamente as escolas e os professores. Não chega dizer que as escolas têm autonomia nesta ou naquela área e, como tal, devem resolver os problemas.

A questão é muito simples. Por exemplo, no ensino básico, há um currículo nacional orientado por competências. Sabe-se que a consistência da avaliação interna é débil e que, sendo assim, podemos pôr em causa a equidade, pois não sabemos se os alunos têm todos acesso ao mesmo tipo de oportunidades de aprendizagem. Não podemos garantir que todos os alunos tiveram oportunidade para aprender este ou aquele domínio essencial e estruturante do currículo. O que faz o Ministério da Educação perante esta situação? Como vai trabalhar com as escolas? Que tipo de meios de formação vai disponibilizar? Vai delinear um plano integrado, coerente e participado para melhorar a qualidade do ensino e da avaliação no ensino básico? Ou pensará que estes problemas se resolvem por si mesmos ou através da mera administração de exames?

A avaliação externa tem obedecido mais a uma espécie de calendário marcado pela força das circunstâncias do que a uma acção deliberadamente planeada e articulada com objectivos e com consequências formativas para o sistema, as escolas, os alunos e os pais.

No que se refere aos exames nacionais do ensino secundário, pode dizer-se que, após um período, difícil, de consolidação relativa a um vastíssimo conjunto de procedimentos que ia desde a concepção e elaboração das provas e da distribuição de papel de prova até aos mecanismos de controlo e segurança do processo, é tempo de se entrar numa fase de desenvolvimento da qualidade.

Sendo os exames uma avaliação externa em larga escala, referida à norma, é necessário garantir que as suas características psicométricas sejam aceitáveis e de acordo com o que são os *standards* internacionalmente recomendados e aceites. E esta questão é relevante porque dela depende a equidade do processo. Ninguém, com certeza, aceitará que o aluno A ou a aluna B tiveram, ou não, acesso ao ensino superior por razões devidas à validade ou fiabilidade dos exames e não por razões devidas às aprendizagens de que eram, ou não, possuidores. Parece-me, por isso, que as entidades responsáveis devem trabalhar no sentido de divulgar esses e outros parâmetros psicométricos tais como os índices de dificuldade e de discriminação dos itens. Sente-se que o processo de exames, que é muito transparente em muitos outros aspectos, deve ser mais transparente e mais aberto à discussão e ao escrutínio dos cidadãos nas questões relativas à qualidade das provas pois, dessa forma, poder-se-á contribuir para uma discussão pública melhor e mais fundamentada. Por isso mesmo, se estranha o facto de não serem amplamente divulgados relatórios em que se analisem e interpretem os dados, explicitando tendências, já que as séries de dados existentes devem permitir fazê-lo.

Outro problema que tem que ser tornado mais transparente é o processo de correcção das provas. É uma matéria da maior importância com efeitos que podem questionar completamente a equidade e a justeza de todo o processo.

Por outro lado, seria importante que as análises e interpretações dos resultados permitissem também caracterizar o currículo que está efectivamente a ser avaliado e que tipo de efeitos é que se sentem, inclusivamente, ao nível das aprendizagens dos alunos. Uma questão sempre muito interessante é a de saber se os alunos estão a aprender mais e melhor após os exames. Ou será que estão a aprender menos? Será que, neste momento, teremos respostas para estas e outras questões tão essenciais? Julgo que o processo de exames do ensino secundário deve ter associado um programa permanente de investigação. Só dessa forma o poderemos tornar mais transparente e levar a desempenhar mais adequadamente as suas funções de avaliação dos alunos e do sistema educativo.

O texto mostra que o país tem lidado mal com as provas de aferição. Neste momento, praticamente ninguém parece levá-las a sério. Nem o próprio Ministério da Educação quando demora cerca de dois anos para divulgar os seus resultados. É um sinal muito negativo que se dá às escolas, aos professores, aos alunos e à sociedade em geral. Refira-se que em muitos países da Europa e da América do Norte existem provas desta natureza que têm um grande impacto na definição de políticas, no desenvolvimento do currículo ou na melhoria das práticas educativas a vários níveis. As provas de aferição têm várias vantagens que não devem ser desprezadas. Podem avaliar domínios do currículo que não o podem ser pelos exames. Podem ser realizadas em disciplinas que não são objecto de exame. Podem servir para dar sinais importantes aos professores e aos alunos acerca do que é importante valorizar. Por outro lado, podem ser administradas pelos próprios professores dos alunos nas suas aulas, o que lhes confere uma natureza menos formal que os exames, permitindo assim gerir o tempo e os procedimentos de administração de forma mais flexível e menos standardizada do que se impõe nos exames. Por exemplo, seria muito importante que se desenvolvessem provas na área das ciências experimentais ou provas com componente de expressão verbal na língua portuguesa e nas línguas estrangeiras. Repare-se que as provas de aferição deverão, preferencialmente, ser administradas a uma amostra de alunos do ano em causa e, por isso, há muito mais facilidade em avaliar estas «áreas» do que num exame.

As provas aferidas têm que ser credibilizadas, têm que ter um plano de administrações e de disciplinas a abranger. É um desperdício administrá-las anualmente. Nenhum país parece estar a fazê-lo. Precisa-se de uma ideia para as provas aferidas que as integre num todo coerente de contribuições para a avaliação dos alunos, das escolas e do sistema. Mais uma vez, dificilmente se poderá conceber um programa de provas desta natureza sem um projecto associado de investigação que descreva, analise e interprete os dados e que seja capaz de intervir no sentido de fazer recomendações que façam sentido para os decisores políticos, para as escolas, para os professores e para os alunos

No domínio das provas internacionais, a situação não é muito diferente. Vamos participando mas sem estratégia, sem objectivos aparentes para o desenvolvimento do sistema, sem ampla discussão e sem elaboração de relatórios nacionais com real impacto na sociedade e, sobretudo, nos níveis de decisão política, nas escolas e nos professores. É importante promover programas de investigação que lhes estejam associados para que nos apropriemos de desenvolvimentos ao nível da elaboração e da análise de itens e de resultados que normalmente aparecem associados a estes projectos e que podem ser-nos úteis para as provas externas.

É preciso definir uma política de avaliação que integre e relacione todas estas componentes, que credibilize e valorize efectivamente a avaliação interna e que promova a sua articulação com uma avaliação externa que pode e deve ter o seu papel no sistema educativo.

REFERÊNCIAS

- Airasian, P. e Abrams, L. (2003). Classroom student evaluation. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 533-548. Dordrecht: Kluwer.
- Alaiz, V. (1993, Junho). School-based and external assessment in Portuguese primary and secondary education. Comunicação apresentada na *19th Conference of the International Association for Educational Assessment* (IAEA). Grand Baie. Ilha Maurícia.
- Alves, J. Carlos. (1997). *Representações da avaliação por parte dos professores dos 2.º e 3.º ciclos em três escolas do distrito de Setúbal*. Tese de mestrado em ciências da educação (Avaliação em educação) não publicada. Universidade Católica Portuguesa. Faculdade de Ciências Humanas.
- Anderson, J. e Bachor, D. (1998). A Canadian perspective on portfolio use in student assessment. *Assessment in Education*, 5, 353-380.
- Antunes, Roque. (1995). *Concepções de alunos do 11.º ano em relação à disciplina de Filosofia*. Tese de mestrado em ciências da educação (Desenvolvimento Pessoal e Social) não publicada. Universidade Católica Portuguesa. Faculdade de Ciências Humanas.
- Archbald, D. e Newmann, F. (1992). Approaches to assessing academic achievement. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 139-180. Albany, NY: State University of New York Press.
- Beaton, A. (1997). The national assessment of educational progress. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, adjustment, and achievement*, pp. 518-530. New York: Academic Press.
- Beaton, A., Postlethwaite, T., Ross, K., Spearritt, D., e Wolf, R. (2000). *The benefits and limitations of international educational achievement studies*. Paris: International Institute for Educational Planning/International Academy of Education.
- Bell, A., Burkhardt, H. e Swan, M. (1992a). Balanced assessment of mathematical performance. In R. Lesh e S. Lamon (Eds.), *Assessment of authentic performance in school mathematics*, pp.119-144. Washington, DC: AAAS.
- Bell, A., Burkhardt, H. e Swan, M. (1992b). Assessment of extended tasks. In R. Lesh e S. Lamon (Eds.), *Assessment of authentic performance in school mathematics*, pp.145-176. Washington, DC: AAAS.
- Bell, A., Burkhardt, H. e Swan, M. (1992c). Moving the system: The contributions of assessment. In R. Lesh e S. Lamon (Eds.), *Assessment of authentic performance in school mathematics*, pp.177-194. Washington, DC: AAAS.
- Benavente, A., Alaiz, V., Barbosa, J., Campos, C., Carvalho, A. e Neves, A. (1995). *Novo modelo de avaliação no ensino básico: Formas de implementação local*. Lisboa: IIE.
- Berlak, H. (1992a). The need for a new science of assessment. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 1-22. Albany, NY: State University of New York Press.
- Berlak, H. (1992b). Toward the development of a new science of educational testing and assessment. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 181-206. Albany, NY: State University of New York Press.
- Berlak, H., Newmann, F., Adams, E., Archbald, D., Burgess, T., Raven, J. e Romberg, T. (Eds.) (1992), *Toward a new science of educational testing and assessment*. Albany, NY: State University of New York Press.
- Bloom, B. (1956). *Taxonomy of educational objectives*. David McKay.
- Boavida, Joaquim. (1996). *Concepções e práticas de avaliação das aprendizagens de professoras do 1.º ciclo do ensino básico: Três estudos de caso*. Tese de mestrado em ciências da educação (Avaliação em educação) não publicada. Universidade Católica Portuguesa. Faculdade de Ciências Humanas.
- Broadfoot, P. (1994). Les résultats de l'enseignement. In OCDE (Ed.), *Évaluer l'enseignement: De l'utilité des indicateurs internationaux*, pp. 260-286. Paris: OCDE.
- California Assessment Program, (1989). *A question of thinking: A first look of students' performance on open-ended questions in mathematics*. Sacramento, CA: California State Department of Education.
- Campos, Cristina. (1996). *Concepções e práticas de professores sobre avaliação das aprendizagens: Dois estudos de caso*. Tese de mestrado em ciências da educação (Avaliação em educação) não publicada. Universidade Católica Portuguesa. Faculdade de Ciências Humanas.
- Chambers, D. (1993). Integrating assessment and instruction. In N. Webb e A. Coxford (Eds.), *Assessment in the mathematics classroom*, pp. 17-25. Reston, VA: NCTM.
- Cockcroft, W. (1982). *Mathematics counts*. Londres: HMSO.
- Conceição, J. (1993a). Ainda o novo sistema de avaliação. *Noesis*, 25, 64-67.
- Conceição, J. (1993b, Junho). External assessment in Portuguese compulsory education. Comunicação apresentada na *19th Conference of the International Association for Educational Assessment* (IAEA). Grand Baie. Ilha Maurícia.
- Conceição, J., Neves, A., Campos, C., Fernandes, D. e Alaiz (sem data). Testes: Sim ou não? In IIE (Ed.), *Pensar avaliação, melhorar a aprendizagem* (Folha B/8). Lisboa: IIE.
- Conselho Nacional de Educação (2000). Parecer n.º 3/2000. *Diário da República*, 180, 13016-13024. II Série.
- Fernandes, D. (1991). *Resolução de problemas e avaliação*. Actas do 2.º Encontro Nacional de Didáticas e Metodologias de Ensino. Aveiro: Universidade de Aveiro.
- Fernandes, D. (1992a). Resolução de problemas: Investigação, ensino, avaliação e formação de professores. In M. Brown, D. Fernandes, J. F. Matos e J. P. Ponte (Eds.), *Educação matemática: Temas de investigação*, pp. 45-104. Lisboa: Instituto de Inovação Educacional.
- Fernandes, D. (1992b). O tempo da avaliação. *Noesis*, 23, 18-21.

- Fernandes, D. (1992c). *Práticas e perspectivas de avaliação: Dois anos de experiência no Instituto de Inovação Educacional*. Documento policopiado não publicado.
- Fernandes, D. (1993a). Complexidade, tensões e mudança na avaliação das aprendizagens. In L. Almeida, J. Fernandes e A. Mourão (Orgs.), *Ensino-aprendizagem da matemática: Recuperação de alunos com baixo desempenho*, pp. 43-60. Riba d'Ave: Didáxis.
- Fernandes, D. (1993b). Some notes on students' assessment in the context of the Portuguese educational reform. Trabalho não publicado apresentado na conferência *Research into pupil assessment and the role of final examinations in secondary education*. Conselho da Europa. Universidade de Jyväskylä: Finlândia.
- Fernandes, D. (1994a). Avaliação das aprendizagens: Das prioridades de investigação e de formação às práticas na sala de aula. *Revista de Educação*, 8, 15-20.
- Fernandes, D. (1994b). Evaluating the educational reform: Viewpoints from a Portuguese experience. In Mauritius Examination Syndicate (Ed.), *1993 IAEA Conference: School-based and external assessments*, pp. 251-260. Reduit, Mauritius: Mauritius Examination Syndicate.
- Fernandes, D. (1994c). Contornos de uma experiência de avaliação desenvolvida no Instituto de Inovação Educacional (1990-1993). *Boletim da Sociedade Portuguesa de Educação Física*, 10/11, 7-32.
- Fernandes, D. (1997). Avaliação na escola básica obrigatória: Fundamentos para uma mudança de práticas. In Pedro da Cunha (Org.), *Educação em debate*, pp. 275-294 Lisboa, Universidade Católica Portuguesa.
- Fernandes, D., Neves, A., Campos, C. e Landa, J. (1996). *Das concepções, práticas e organização da avaliação das aprendizagens à formação de professores*. (Relatório do 1.º ano do Projecto PI/12/94 financiado pelo Instituto de Inovação Educacional.) Documento policopiado não publicado.
- Fernandes, D., Ramalho, G. e Lemos, V. (1991). *Opiniões dos professores dos ensinos básico e secundário relativamente às medidas constantes no projecto do sistema de avaliação dos alunos*. Lisboa: IIE.
- Foucault, M. (1979). *Discipline and punish: The birth of prison*. New York: Vantage.
- Gardner, H. (1983). *Frames of mind*. New York: Basic Books.
- Gardner, H. (1991). *The unschooled mind: How children think and how schools should teach*. New York: Basic Books.
- Gardner, H. e Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher*, 18(8), 4-9.
- Gifford, B. e O' Connor (Eds.) (1992). *Changing assessments: Alternative views of aptitude, achievement and instruction*. Dordrecht: Kluwer.
- Gil, Dulcinea. (1997). *Reflexões de professores do 2.º ciclo do ensino básico sobre avaliação das aprendizagens. Tese de mestrado em ciências da educação* (Avaliação em educação) não publicada. Universidade Católica Portuguesa. Faculdade de Ciências Humanas.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. Londres: Falmer.
- Gipps, C. e Stobart, G. (2003). Alternative assessment. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 549-576. Dordrecht: Kluwer.
- Goldstein, H. (1996). International comparisons of student achievement. In A. Little e A. Wolf (Eds.), *Assessment in transition: Learning, monitoring and selection in international perspective*, pp. 58-87. Oxford: Pergamon.
- Guba, E. e Lincoln, Y. (1989). *Fourth generation evaluation*. London: Sage.
- Guba, E. e Lincoln, Y. (1994). Competing paradigms in qualitative research. In N. Denzin e Y. Lincoln (Eds.), *Handbook of qualitative research*, pp. 105-117. Londres: Sage.
- Instituto de Inovação Educacional (sem data). *Pensar a avaliação, melhorar a aprendizagem*. Lisboa: Autor.
- Instituto de Inovação Educacional (1992a). *Estudo comparativo dos sistemas de avaliação dos alunos em quatro países europeus*. Lisboa: Autor.
- Instituto de Inovação Educacional (1992b). *Avaliar é aprender*. Lisboa: Autor.
- Johnson, R. (2003). The development and use of school profiles. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 827-842. Dordrecht: Kluwer.
- Joint Committee on Standards for Educational Evaluation (1981). *Principles and by-laws*. Kalamazoo: MI: Western Michigan University Evaluation Center.
- Jones, L. (2003). National assessment in the United States: The evolution of a Nation's report card. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 883-904. Dordrecht: Kluwer.
- Keeves, J. (1995). *The world of school learning: Selected key findings from 35 years of IEA research*. Amsterdam: IEA Secretariat.
- Kellaghan, T. (1996). IEA studies and educational policy. *Assessment in Education*, 3, 143-160.
- Kellaghan, T. (2003). Local, national, and international levels of system evaluation. Introduction. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 873-882. Dordrecht: Kluwer.
- Kellaghan, T. e Grisay, A. (1995). International comparisons of student achievement: Problems and prospects. In OECD (Ed.), *Measuring what students learn*, pp. 41-61. Paris: OECD.
- Kellaghan, T. e Madaus, G. (2000). Outcome evaluation. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), pp. 97-112. Dordrecht: Kluwer.
- Kellaghan, T. e Madaus, G. (2003). External (public) examinations. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 577-602. Dordrecht: Kluwer.

- Kellaghan, T. e Stufflebeam, D. (Eds.) (2003). *International handbook of educational evaluation*. Dordrecht: Kluwer.
- Kilpatrick, J. (1992). Some issues in the assessment of mathematical problem solving. In J. Ponte, J. F. Matos, J. M. Matos e D. Fernandes (Eds.), *Mathematical problem solving and new information technologies: Research in contexts of practice*, pp. 37-44. Berlim: Springer-Verlag.
- Koretz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education*, 5, 309-334.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Lange, J. de (1987). *Mathematics, insight and meaning*. Utrecht, Holanda: OweOC.
- Lange, J. de (1993). Assessment in problem-oriented curricula. In N. Webb e A. Coxford (Eds.), *Assessment in the mathematics classroom*, pp. 197-208. Reston, VA: NCTM.
- Lapointe, A. (1986). Testing in the USA. In D. Nuttall (Ed.), *Assessing educational achievement*, pp. 114-124. London: Falmer.
- Lemos, V. (1992). Novo sistema de avaliação: Enquadramento de uma mudança. Comunicação apresentada no seminário *O Novo Sistema de Avaliação* Instituto de Inovação Educacional: Lisboa.
- Lemos, V. (1993, Junho). Attitudes on assessment and the in-service training of teachers. Comunicação apresentada na *19th Conference of the International Association for Educational Assessment (IAEA)*. Grand Baie. Ilha Maurícia.
- Lemos, V., Neves, A., Campos, C., Conceição, J. e Alaiz, V. (1992). *A nova avaliação da aprendizagem: O direito ao sucesso*. Lisboa: Texto Editora.
- Lester, F. e Kroll, D. (1990). Assessing student growth in mathematical problem solving. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics*, pp. 53-70. Washington: AAAS.
- Lobo, Aldina (1998). A. A. A. (*Aprendizagem Assistida pela Avaliação*): *Um sorriso difícil sobre o novo sistema de avaliação do ensino básico*. Porto: Porto Editora.
- Madaus, G. e Kellaghan, T. (2000). Models, metaphors, and definitions in evaluation. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), pp. 19-32. Dordrecht: Kluwer.
- Madaus, G. e Stufflebeam, D. (2000). Program evaluation: A historical overview. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), pp. 3-18. Dordrecht: Kluwer.
- Madaus, G., Haney, W. e Kreitzer, A. (2000). The role of testing in evaluations. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), pp. 113-126. Dordrecht: Kluwer.
- Ministério da Educação (1991). Despacho normativo n.º 162. *Diário da República*, 244, 10598-10601. I Série B.
- Ministério da Educação (1992). Despacho normativo n.º 98-A. *Diário da República*, 140, 2908(2)-2908(4). I Série B.
- Ministério da Educação (1993). Despacho normativo n.º 338. *Diário da República*, 247, 5934-5937. I Série B.
- Ministério da Educação (1994). Despacho normativo n.º 644-A. *Diário da República*, 214, 5556 (2). I Série B.
- Ministério da Educação (2000). Despacho n.º 5437. *Diário da República*, 58, 4613. II Série B.
- Ministério da Educação (2000). *Provas de aferição do ensino básico: 4.º ano – 2000*. Lisboa: Autor.
- Ministério da Educação (2001). Decreto-Lei n.º 6. *Diário da República*, 15, 258-265. I Série A.
- Ministério da Educação (2001). Despacho normativo n.º 30. *Diário da República*, 166, 4438-4441. I Série B.
- Ministério da Educação (2002). *Provas de aferição do ensino básico: 4.º e 6.º anos – 2001*. Lisboa: Autor.
- Ministério da Educação (2002). Despacho n.º 2205. *Diário da República*, 23, 1762. II Série.
- Ministério da Educação (2002). Decreto-Lei n.º 209. *Diário da República*, 240, 6807-6810. I Série A.
- Ministério da Educação (2002). Lei n.º 31. *Diário da República*, 294, 7952-7953. I Série A.
- Ministério da Educação (2003). Despacho n.º 474. *Diário da República*, 8, 376. II Série.
- Ministério da Educação (2004). Despacho n.º 1911. *Diário da República*, 23, 1489. II Série.
- Mislevy, R., Wilson, M., Ercikan, K. e Chudowsky (2003). Psychometric principles in student assessment. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 489-532. Dordrecht: Kluwer.
- National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Autor.
- National Council of Teachers of Mathematics (1991). *Professional standards for teaching mathematics*. Reston, VA: Autor.
- National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics*. Reston, VA: Autor.
- National Research Council (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.
- National Research Council (1993). *Measuring up: Prototypes for mathematics assessment*. Washington, DC: National Academy Press.
- Neves, Anabela. (1996). *Observação nas concepções e práticas de dois professores do ensino básico*. Tese de mestrado em ciências da educação (Avaliação em educação) não publicada. Universidade Católica Portuguesa. Faculdade de Ciências Humanas.
- Nevo, D. (1986). The conceptualization of educational evaluation: An analytical review of the literature. In E. R. House (Ed.), *New directions in educational evaluation*, pp. 15-29. London: Falmer.
- Newmann, F. e Archbald, D. (1992). The nature of authentic academic achievement. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 71-84. New York: SUNY.
- Nuttall, D. (Ed.) (1986). *Assessing educational achievement*. London: Falmer.
- O'Connor, M. (1992). Rethinking aptitude, achievement, and instruction: Cognitive science research and the framing of assessment policy. In B. Gifford e M. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (p. 9-36). Norwell, MA: Kluwer.

- OECD (1997). *Education at a glance: OECD indicators*. Paris: Autor.
- OECD (2000). *Education at a glance: OECD indicators*. Paris: Autor.
- OECD (2003). The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills. Paris: Autor.
- Oliveira, I., Pereira, J. e Fernandes, D. (1993). *Desenvolvimento de instrumentos de avaliação da aprendizagem em Matemática*. Lisboa: IIE.
- Oliveira, I., Pereira, J. e Fernandes, D. (1994). *Seis propostas de avaliação: Matemática*. Lisboa: IIE.
- Pearlman, M. e Tannembaum, R. (2003). Teacher evaluation practices in the accountability era. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 609-642. Dordrecht: Kluwer.
- Phye, G. (Ed.) (1997). *Handbook of classroom assessment: Learning, adjustment, and achievement*. New York: Academic Press.
- Plomp, T., Howie, S. e McGaw, B. (2003). International studies of educational achievement. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 951-978. Dordrecht: Kluwer.
- Raven, J. (1992). A model of competence, motivation, and behavior, and a paradigm of assessment. In H. Berlak, F. Newmann, E. Adams, D. Archbald, T. Burgess, J. Raven e T. Romberg (Eds.), *Toward a new science of educational testing and assessment*, pp. 85-116. Albany, NY: State University of New York Press.
- Richardson, V. (Ed.) (2001). *Handbook of research on teaching* (4th Edition). American Educational Research Association. New York: Macmillan.
- Robitaille, D., Beaton, A. e Plomp, T. (2000). *The impact of TIMSS on the teaching and learning of mathematics and science*. Vancouver: Pacific Educational Press.
- Romberg, T., Zarinnia, E. e Collis, K. (1990). A new world view of assessment in mathematics. In G. Kulm (Ed.), *Assessing higher order thinking in mathematics*, pp. 21-38. Washington, DC: AAAS.
- Sanders, J. e Davidson, E. (2003). A model for school evaluation. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 807-826. Dordrecht: Kluwer.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic Press.
- Scriven, M. (2000). Evaluation ideologies. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), pp. 249-278. Dordrecht: Kluwer.
- Sebring, P. e Boruch, R. (1991). How the national assessment of educational progress is used: An update of an exploratory study. In R. Stake (Ed.), *Advances in program evaluation: Effects of mandated assessment on teaching* (Volume 1, Part B), pp. 255-276. London: Jay Press.
- Shepard, L. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of research on teaching* (4th Edition). American Educational Research Association. New York: Macmillan.
- Sikula, J. (Ed.) (1996). *Handbook of research on teacher education* (2nd Edition). Association of Teacher Educators. New York: Macmillan.
- Stake, R. (Ed.) (1991a). *Advances in program evaluation: Using assessment policy to reform education* (Volume 1, Part A). London: Jay Press.
- Stake, R. (Ed.) (1991b). *Advances in program evaluation: Effects of mandated assessment on teaching* (Volume 1, Part B). London: Jay Press.
- Stake, R. (2000). Program evaluation, particularly responsive evaluation. In D. Stufflebeam, G. Madaus e T. Kellaghan, (Eds.) *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition), pp. 343-362. Dordrecht: Kluwer.
- Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessment. *Assessment in Education*, 5, 335-352.
- Stiggins, R. e Conklin, N. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Stufflebeam, D. (2003). Institutionalizing evaluation in schools. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 775-806. Dordrecht: Kluwer.
- Stufflebeam, D., Madaus, G. e Kellaghan, T. (Eds.) (2000). *Evaluation models: Viewpoints on educational and human services evaluation* (2nd Edition). Dordrecht: Kluwer.
- Tellez, K. (1996). Authentic assessment. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd Edition), pp. 704-721. Association of Teacher Educators. New York: Macmillan.
- The Vermont Department of Education (1991). *Looking beyond «The Answer»: The report of Vermont's mathematics portfolio assessment program* (Pilot year, 1990-1991). Vermont: Autor.
- Torrance, H. (2003). Assessment of the national curriculum in England. In T. Kellaghan e D. Stufflebeam (Eds.), *International handbook of educational evaluation*, pp. 905-928. Dordrecht: Kluwer.
- Tunstall, P. e Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22, 389-404.
- Valencia, S., Hiebert, E. e Afflerbach, P. (1994). *Authentic reading assessment: Practices and possibilities*. Newark: International Reading Association.
- Webb, N. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning*, pp. 661-686. New York: MacMillan.
- Webb, N. (1993). Assessment for the mathematics classroom. In N. Webb e A. Coxford (Eds.), *Assessment in the mathematics classroom*, pp. 1-6. Reston, VA: NCTM.
- Webb, N. e Coxford, A. (Eds.) (1993). *Assessment in the mathematics classroom*. Reston, VA: NCTM.



©Texto Editora 2004

ATENDIMENTO A PROFESSORES

Livraria Cacém Edifício Texto Editora - Estrada de Paço de Arcos, 66, 2735-336 Cacém

Livraria Lisboa Rua Joaquim Paço D'Arcos, 13, 1500-365 Lisboa

Livraria Porto Rua Damião de Góis, 45, 4050-225 Porto

Linha do Professor 707 231 231

Fax 21 427 22 09

E-mail: escolar@te.pt